

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:14:15 ; Search time 30 Seconds
(without alignments)
2503.232 Million cell updates/sec

Title: US-09-914-042-1
Perfect score: 5273
Sequence: 1 MPDQISVEFVAETHEDYK.....DGPGRKGAFFPVFVHFAD 1006

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/1aa/5A COMB.pap:*
2: /cgn2_6/ptodata/1/1aa/5B COMB.pap:*
3: /cgn2_6/ptodata/1/1aa/6A COMB.pap:*
4: /cgn2_6/ptodata/1/1aa/6B COMB.pap:*
5: /cgn2_6/ptodata/1/1aa/PTUS COMB.pap:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5253	99.6	1006	US-09-023-905A-12	Sequence 12, Appli
2	3874	73.5	982	US-09-023-905A-7	Sequence 7, Appli
3	3155	59.8	903	US-09-023-905A-10	Sequence 10, Appli
4	3013.5	57.1	1129	US-09-023-905A-2	Sequence 2, Appli
5	2954.5	56.0	1151	US-09-023-905A-4	Sequence 4, Appli
6	1387	26.3	307	US-09-023-905A-36	Sequence 36, Appli
7	484.5	9.2	134	US-09-023-905A-30	Sequence 30, Appli
8	453	8.6	166	US-09-270-767-33258	Sequence 33258, A
9	419	7.9	605	US-09-828-303-19	Sequence 19, Appli
10	347.5	6.6	225	US-09-270-767-42748	Sequence 42748, A
11	278	5.3	802	US-09-949-016-11166	Sequence 6235, Ap
12	262	5.0	750	US-09-949-016-11166	Sequence 11166, A
13	261.5	5.0	913	US-09-248-796A-15734	Sequence 15734, A
14	223	4.2	377	US-09-270-767-42077	Sequence 42077, A
15	217	4.1	1255	US-09-080-897-4	Sequence 4, Appli
16	217	4.1	1255	US-08-899-595-1	Sequence 1, Appli
17	217	4.1	1255	US-09-323-735-4	Sequence 4, Appli
18	214.5	4.1	506	US-09-949-016-11282	Sequence 11282, A
19	209.5	4.0	503	US-09-598-287A-2	Sequence 2, Appli
20	208	3.9	507	US-09-599-287A-24	Sequence 24, Appli
21	204.5	3.9	2321	US-09-230-652-2	Sequence 2, Appli
22	201	3.8	581	US-09-949-016-9978	Sequence 9978, Ap
23	199	3.8	802	US-09-823-240A-2	Sequence 2, Appli
24	198	3.8	709	US-09-949-016-6809	Sequence 6809, Ap
25	198	3.8	728	US-09-949-016-7213	Sequence 7213, Ap
26	195	3.7	62	US-08-687-702-18	Sequence 18, Appli
27	195	3.7	1312	US-09-345-882-29	Sequence 29, Appli

ALIGNMENTS

RESULT 1
US-09-023-905A-12
; Sequence 12, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023.905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-023-905A-12

Query Match	99.6%	Score 5253;	DB 4;	Length 1006;
Best Local Similarity	99.7%	Pred. No. 0;		
Matches 1003;	Conservative	0;	Mismatches	3;
Indels	0;	Gaps	0;	
Qy	1	MPDQISVEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMLVKMKKSVA	60	
Db	1	MPDQISVEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMLVKMKKSVA	60	
Qy	61	INSSGLAHVNEEQYTOALEKFGGNCVCRDDPDGLGSAPLKFVSFTKLTALFNLIQNNN	120	
Db	61	INSSGLAHVNEEQYTOALEKFGGNCVCRDDPDGLGSAPLKFVSFTKLTALFNLIQNNN	120	
Qy	121	NIISFPDLSLLKGLDKGVKGLKKPFDKAWKDYETKITKIEKKEKHAHLHGMIRTEISG	180	
Db	121	NIISFPDLSLLKGLDKGVKGLKKPFDKAWKDYETKITKIEKKEKHAHLHGMIRTEISG	180	
Qy	181	AEIAEEMEKERRFQLOMCEYLLKVNKEIKIKGVDLLQNLIKYFHAQCNFFODGLKAVES	240	
Db	181	AEIAEEMEKERRFQLOMCEYLLKVNKEIKIKGVDLLQNLIKYFHAQCNFFODGLKAVES	240	
Qy	241	LKPSIETLSTDLHTIKQAQDEERQLIQLRDLKLSALQVEOKESQIRQSTAYSJLHQPG	300	
Db	241	LKPSIETLSTDLHTIKQAQDEERQLIQLRDLKLSALQVEOKESQIRQSTAYSJLHQPG	300	

Same as W0


```

RESULT 3
US-09-023-905A-10
; Sequence 10, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Danio rerio
US-09-023-905A-10

Query Match      59.8%; Score 3155; DB 4; Length 903;
Best Local Similarity 65.3%; Pred. No. 2.6e-226;
Matches 621; Conservative 98; Mismatches 174; Indels 59; Gaps 10;

Qy      1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMLVYMKKSVKA 60
Db      1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMLVYMKKSVKA 60
Qy      61 INSSGLAHVNEEQYTOALEKFGGNCVCRDDPLGSAFLKFSVFTKELTALPKNIQNN 120
Db      61 INSSGLAHVNEEQYTOALEKFGGNCVCRDDPLGSAFLKFSVFTKELTALPKNIQNN 120
Qy      121 NIISPLDLSLLKGDLLKVGKDLKPPDKAWKDYETKIKIEKEKEHAKLHGMIRTEISG 180
Db      121 NIISPLDLSLLKGDLLKVGKDLKPPDKAWKDYETKIKIEKEKEHAKLHGMIRTEISG 180
Qy      181 AEIAEMEKERRFLQMCVEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKAVES 240
Db      181 AEIAEMEKERRFLQMCVEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKAVES 240
Qy      241 LKPSIETLSTDLHTTIKQADEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSLSHQPG 300
Db      241 LKPSIETLSTDLHTTIKQADEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSLSHQPG 300
Qy      301 NKEHGTENGSLYKSDGIRKWKQKSVKNGFLTISHGTANRPPAKNLNLTTCQVKNP 360
Db      301 NKEHGTENGSLYKSDGIRKWKQKSVKNGFLTISHGTANRPPAKNLNLTTCQVKNP 360
Qy      361 BEKKCFDLISHDRTYHFOADESOECQIWMVSLQNSKEBALNNAFKGDDNTGNNIVOEIT 420
Db      361 BEKKCFDLISHDRTYHFOADESOECQIWMVSLQNSKEBALNNAFKGDDNTGNNIVOEIT 420
Qy      421 KEIIEVQRMGNDVCCDGPDPPTLSTNLGILTCIECSGIRHRELGVHYSMPQSLTLDV 480
Db      421 KEIIEVQRMGNDVCCDGPDPPTLSTNLGILTCIECSGIRHRELGVHYSMPQSLTLDV 480
Qy      481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPGSDNNARKDYITAKYIERRYARKKH 540
Db      481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPGSDNNARKDYITAKYIERRYARKKH 540
Qy      541 ADNAKHLHSLCAVKTTRDIFGLIQAYADGVDTETKIPLANGHEPDETALHLAVRSVDRFS 600
Db      541 ADNAKHLHSLCAVKTTRDIFGLIQAYADGVDTETKIPLANGHEPDETALHLAVRSVDRFS 600
Qy      599 ADNAARLHALCAVKSRRDIFPSLIQVYABGLDMETINOPNQHGETSLHLAVRMVDRNS 599
Db      599 ADNAARLHALCAVKSRRDIFPSLIQVYABGLDMETINOPNQHGETSLHLAVRMVDRNS 599
Qy      601 LHIVDFLVQNSGNDLKQTKGKTGKTALHYCCCLTDNSECMLKLLRGKASVITNDAGETALD 659
Db      601 LHIVDFLVQNSGNDLKQTKGKTGKTALHYCCCLTDNSECMLKLLRGKASVITNDAGETALD 659
Qy      661 AKRLKHECEBELLTOALSGRNSVHVYEWRLHEDLDESDDDMDKELQPSSENRDRP 720
Db      661 AKRLKHECEBELLTOALSGRNSVHVYEWRLHEDLDESDDDMDKELQPSSENRDRP 720
Qy      660 AORLKHSKEBELLTOAQTKGFNVHVHVDYDRLHNEDEDESEDEK--PIPIRREBP 717
Db      660 AORLKHSKEBELLTOAQTKGFNVHVHVDYDRLHNEDEDESEDEK--PIPIRREBP 717
Qy      721 ISFYQLGNSQNAVSLARDAANLAKQKRAFMPISILQNETYTGALLSGSPPAQPAAPS 780
Db      721 ISFYQLGNSQNAVSLARDAANLAKQKRAFMPISILQNETYTGALLSGSPPAQPAAPS 780
Qy      718 ISCVPGSGPMNMSALARDVANVNNKQRAFIPSMNMNITYGTMLDPNSPL--GLPG 775
Db      718 ISCVPGSGPMNMSALARDVANVNNKQRAFIPSMNMNITYGTMLDPNSPL--GLPG 775
Qy      781 TTSAPPLPPRNVGKQVQTASSANTLWKTNSVSDGSRQSSSDPPPAVHPPPLPLAVTSTN 840
Db      781 TTSAPPLPPRNVGKQVQTASSANTLWKTNSVSDGSRQSSSDPPPAVHPPPLPLAVTSTN 840
Qy      776 VPGIPLPPPLRGRG-----W---SPMENIGRQSCSDPA---NPQTPEQNNSVY 820
Db      776 VPGIPLPPPLRGRG-----W---SPMENIGRQSCSDPA---NPQTPEQNNSVY 820
Qy      841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPLPPQPPSRLPQKPKPAGTDKST 900
Db      841 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPLPPQPPSRLPQKPKPAGTDKST 900
Qy      821 VLPAPAPPPPA-----PKRPPPPDKASLLPAAATAP-----PAP-----SA 857
Db      821 VLPAPAPPPPA-----PKRPPPPDKASLLPAAATAP-----PAP-----SA 857
Qy      901 PLTNKGQPRGPDLSATEALGPLSNAMVLPAPPMPRKSOATKLPKRVKA 951
Db      901 PLTNKGQPRGPDLSATEALGPLSNAMVLPAPPMPRKSOATKLPKRVKA 951
Qy      858 PLLPPAPLRPAP-----VPPAPVMPPTSSLTDKVLSLSKA 893
Db      858 PLLPPAPLRPAP-----VPPAPVMPPTSSLTDKVLSLSKA 893

RESULT 4
US-09-023-905A-2
; Sequence 2, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Bos sp.
US-09-023-905A-2

Query Match      57.1%; Score 3013.5; DB 4; Length 1129;
Best Local Similarity 55.3%; Pred. No. 1.3e-215;
Matches 627; Conservative 129; Mismatches 225; Indels 153; Gaps 23;

Qy      1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMLVYMKKSVKA 60
Db      1 MPDQISVSEFVAETHEDYKAPTASFTTRTAQCRNTVAALBEALDVRMLVYMKKSVKA 60
Qy      61 INSSGLAHVNEEQYTOALEKFGGNCVCRDDPLGSAFLKFSVFTKELTALPKNIQNN 120
Db      61 INSSGLAHVNEEQYTOALEKFGGNCVCRDDPLGSAFLKFSVFTKELTALPKNIQNN 120
Qy      81 IYNSGDHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFAVKFSTLTKELTLLKLLQGLS 140
Db      81 IYNSGDHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFAVKFSTLTKELTLLKLLQGLS 140
Qy      121 NIISPLDLSLLKGDLLKVGKDLKPPDKAWKDYETKIKIEKEKEHAKLHGMIRTEISG 180
Db      121 NIISPLDLSLLKGDLLKVGKDLKPPDKAWKDYETKIKIEKEKEHAKLHGMIRTEISG 180
Qy      141 HNVITLDSLLKGDLLKVGKDLKPPDKAWKDYETKIKIEKEKEHAKLHGMIRTEISG 200
Db      141 HNVITLDSLLKGDLLKVGKDLKPPDKAWKDYETKIKIEKEKEHAKLHGMIRTEISG 200
Qy      181 AEIAEMEKERRFLQMCVEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKAVES 240
Db      181 AEIAEMEKERRFLQMCVEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKAVES 240
Qy      201 AEIAEMEKERRFLQMCVEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKADK 260
Db      201 AEIAEMEKERRFLQMCVEYLLKVEIKKGVLDLQNLKLYFHAQCNFFODGLKADK 260
Qy      241 LKPSIETLSTDLHTTIKQADEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSLSHQ 297
Db      241 LKPSIETLSTDLHTTIKQADEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSLSHQ 297
Qy      261 LKQYIEKLAADIYNIKOTQDEBKQLTALDLIKSLQDQKESRRDSQROG-GYSMHQ 319
Db      261 LKQYIEKLAADIYNIKOTQDEBKQLTALDLIKSLQDQKESRRDSQROG-GYSMHQ 319
```

```
Qy 298 POGNKEHCTERNGLYKKS DGIKVMQKRCVKNGELTISHGTANRPPAKNLNLTTCOVK 357
Db 320 LQGNKEYSEKGYLLKKS DGIKVMQKRCVKNGELTISHGTANRPPAKNLNLTTCOVK 379
Qy 358 TNPEKKCFDLISHDRTHYFQAEDQECCQIWMVLSQNSKEEALNNAFAGDDNTGNNIVQ 417
Db 380 PNAEDKKSFDLISHNRTYHFAEDSQDYVAMISVLTNSKEEALTNWAFGEQSGESSL-E 438
Qy 418 ELTKEIIEVQRMGTNDVCCDGPADPTWLTSTNLGILTCIECGIHRBLGVHSPMQSLT 477
Db 439 ELTKAIIEDVQRLPGNDVCCDGSAPETWLTSTNLGILTCIECGIHRBMGVHISRIQSL 498
Qy 478 LDVLGTSSELLAKNIGNAGNEIMECCLPADSVKPNQSDMNAKDYITAKYTERRVAR 537
Db 499 LDKLGTSELLAKNIGNAGNEIMECCLPADSVKPNQSDMNAKDYITAKYTERRVAR 557
Qy 538 KKHADNAKLSLCEAVTRDIFGLQAYADGVLTETIKIPLANGHEPDETALHLAVRSD 597
Db 558 KTCSSSSAKNELLEBAIKSRDILLALIQYVAGVELWEPL-LEPCQELGETALHLAVRTAD 616
Qy 598 RTSLHIVDFVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETP 657
Db 617 QTSLLHVDVFNQCNGLDKQALGNTALHYCSMYSKPECLKLLRSKPTVDVWNAQGETA 676
Qy 658 LDIARLAKHECEELLTQALSGRFSNVHVEYEWRLHEDLDESDDWDEKLOPSENRE 717
Db 677 LDIARLAKATQCEDLLSQAQSGKFNPHVHVEYEWRLHEDLDESDDWDEKLOPSENRE 736
Qy 718 DRPISFYQLGNSQNSVLSARDAANLAKQKRAFMPISILQNETYGA-----LLSGSP 771
Db 737 PRQSFCHSSSISPOD---KLSLPGFSTPRDKQL-----SYGAFTNQIFVSTSTD 784
Qy 772 PPAQAPASTTSAPPLPRNVGK-----VQTASSANTLWKTNSVSVGSGSRQ 819
Db 785 SPTSPGIA---EAPPLPRNATKGGPPGPTLPLSTQSSGSLSKRSPPPPPG-HKR 839
Qy 820 SSSPPPAVHPLPLR--VTSNPLTPTPPPPVAKTTPSVMEALSQ-----862
Db 840 TLSQPPSPPLGHPNKGAVPWNQDVPSSS---SKTTNKFELGSSQSGTSGAKTALVPRV 896
Qy 863 ---PSKPA-----PPGI---SQI-----RPPQ---LPPQ-----883
Db 897 LPKLPQKVALRKTTSHSLDLKANVPPEIQKSSQLTELQKPPPGDLPPKTELA PKP 956
Qy 884 ---PSRLPQK-----KPAAGTDKSTPLTNKGQPR 909
Db 957 PIGDLPPKPGELPPKPOLGDLPPKPOLADPPKPOLPPKPOLGELLAKPQTGDASPK 1016
Qy 910 G-----PVDLS-----ATEALGPLSNAMVLOPPAPMPKSOATKLPKRVKAL 952
Db 1017 AQPPELTPKSHPADLSPNVKQASEDNTDFTLTP-ETPVPPLPKINTGSKVRVRKTI 1075
Qy 953 YNCVADNPDELTFSEGDVIIVDGEQDQEWIGHIDGDPGRKGAFPVSVFHTAD 1006
Db 1076 YDCQADNDELTFMEGEVIVTGEEDQEWIGHIEGQPERKGVFPVSVFHTLSD 1129
```

RESULT 5

```
US-09-023-905A-4
; Sequence 4, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; THEREFOR
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
```

```
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Danio rerio
; US-09-023-905A-4
```

```
Query Match 56.0%; Score 2954.5; DB 4; Length 1151;
Best Local Similarity 54.1%; Pred. No. 3.2e-211;
Matches 622; Conservative 126; Mismatches 239; Indels 163; Gaps 24;

Qy 1 MPDQISVSFVETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
Db 21 MPDQISVSFLETSETDYNSTPTSTTRTQSCRTNVNLEALDQDRTALQKVKKS VKA 80
Qy 61 INSSGLAHVNEEEOYTQALEKFGNGCVCRDDPDGLGSAPLKFPSVFTKELTALFKNLIONMN 120
Db 81 IYNSGQEHVQNEENYGOALDKFGNSFISRDNSDLGTAFIKFSGLIKELAAKLLKLLQSL 140
Qy 121 NIIISFPLDLSLLKGLKGVKGDLPKPPDKAWKDYETKITKIEKKEKHAHGMIRTEISG 180
Db 141 HNVIFTLDSLLKGLKGVKGDLPKPPDKAWKDYETKITKIEKKEKHAHGMIRTEITG 200
Qy 181 AEIAEEMEKERRFPOLQMCYELLKNEIKKGVLDLONLIKYFHAQCNFPQDGLKAVES 240
Db 201 AEIAEEMEKERRIFOLQMCYELLKNEIKKGVLDLONLIKYFHAQCNFPQDGLKATDK 260
Qy 241 LKPSIETLSTLHTIKQAQDEERQLIQLRDILKASALQVEOKESQIRQSTAYSILHOPQ 300
Db 261 LKQYIEKLAADLYNIKQDDEEKQLTALRDLLIKSSQLDQKESQSKS-GYSHWQLQG 319
Qy 301 NKEHGTBRNGLYKKS DGIKVMQKRCVKNGELTISHGTANRPPAKNLNLTTCOVKTNP 360
Db 320 NKEFGSEKKGYLFPKSDGIRKVMQKRCVKNGELTISHGTANRPPAKNLNLTTCOVKPSG 379
Qy 361 EEKCKFDLISHDRTHYFQAEDQECCQIWMVLSQNSKEEALNNAFAGDDNTGNNIVQELT 420
Db 380 EDKCKFDLISHNRTYHFAEDQECCQIWMVLSQNSKEEALNNAFAGDDNTGNNIVQELT 438
Qy 421 KEIISVORMTNDVCCDGPADPTWLTSTNLGILTCIECGIHRBLGVHSPMQSLTLDV 480
Db 439 KAIIEDVLRIPGNEVCCDGPADPTWLTSTNLGILTCIECGIHRBLGVHSPMQSLTLDV 498
Qy 481 LGTSELLAKNIGNAGNEIMECCLPADSVKPNQSDMNAKDYITAKYTERRVAR 540
Db 499 LGTSELLAKNIGNAGNEIMECCLPADSVKPNQSDMNAKDYITAKYTERRVAR 557
Qy 541 ADNAKLSLCEAVTRDIFGLQAYADGVLTETIKIPLANGHEPDETALHLAVRSD 600
Db 558 TTATARQCDLYEAVTRDILMALIQYVAGVELWEPL-LEPCQELGETALHLAVRSD 616
Qy 601 LHIYDVLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETP 660
Db 617 LHLVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETP 676
Qy 661 AKRLKHECEELLTQALSGRFSNVHVEYEWRLHEDLDESDDWDEKLOPSENREDRP 720
Db 677 ARRLNVQCEELLVEAAAGRNPHVHVEYEWRLHEDLDESDDWDEKLOPSENREDRP 736
Qy 721 ISFYQLGNSQNSVLSARDAANLAKQKRAFMPISILQNETYGA-----AFM-----754
Db 737 QSF-----CHSSSVS-PQEKLTLPGLYGHDRDKQRLSYGAFANPVPVSTSTETPASPVS 788
Qy 755 -PSILQNETYCALLSGSP-----PPAQAAPSTTSAPPLP---788
Db 789 GPTTASKTAKAPSCGPTSLPLGSSQSGSSSTLSKKRAPPPPGHKRTHSDPPSPVLQ 848
Qy 789 -PRNVGKVQTASSANTLWKTNSVSVGSGSRQSSSD-----PPAVHPLP-----831
```


Db 20 PSLIAVSEFEVETRSYSSPTTSFASRMPCDRTHTIGVLEERLEFRDREGLTKLAKVAKAI 79
Qy 62 NSSGLAHVENEEOYTOALEKFGGNCVCDDPDLGSAFKFSVFTKELTALFNKLIQNMN 121
Db 80 HNSGNTHVDNEMFVVRALERLGGKVIQDEPDIGAAFLKFSVVTKELSAKMTLMQINN 139
Qy 122 IISPLDLSLLKGLKGVKGLKPKPDK 148
Db 140 IVMFVDSMLKSELKGVAGDMKRPDK 166
RESULT 9
US-09-828-303-19
; Sequence 19, Application US/09828303
; Patent No. 667504
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNET, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUTING
; TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND
; FILE OF INVENTION: METHODS OF USE IN PLANTS
; FILE REFERENCE: 16313-0030
; CURRENT APPLICATION NUMBER: US/09/828,303
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 19
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-09-828-303-19

Query Match 7.9%; Score 419; DB 4; Length 605;
Best Local Similarity 23.9%; Pred. No. 9.9e-23;
Matches 141; Conservative 96; Mismatches 198; Indels 156; Gaps 19;
Qy 211 KKGVDLLQNLKIKYFHAQCNFFODGLKAVESLKPSIETLSTDLHTIKQAQDEERRLIQLR 270
Db 10 KKKFEFLAVSGTMDAHLRYFQGYELLHOMEPYIHQVLT---YAQSKERANYEQAA 66
Qy 271 DILKSALQ-VEQKESQIR-----QSTAYSILHQ-----PQCNKEHGTERNGS 311
Db 67 DRMQEYRQEVERESQSIDFSSSGDGIQGVGRSSHKMIEAVMQSTPKGQIQ--TLKQY 124
Qy 312 LYKKSDBGIRKVMQKCSV-KNGFL-----TISHGTANRPPAKNLNLTQV 356
Db 125 LKRSNTNLGDKWRFFVLDGRGMLYYRKQWKGKPTDEKNVAHT-----VNLITSTI 177
Qy 357 KTNPEE---KKCFDLISHDRTHYFQAEDQECCQIWMVSLQNSKEBALNNAFAGDDNTGEN 413
Db 178 KIDASQSLRFRFCRIISPAKSYTLQAEINAIDRMWDMDKITGVISLLNNQI-SEQVDGED 236
Qy 414 NIV-----QELTEIIESEVORMTGNVCCDGPDPFWLSTNLTGILTCESGIHREL 466
Db 237 SDVSRGASDQSGHERPLDLVRKGNDA CADCGAADPDWASLNLGILLCTECSGVHRM 296
Qy 467 GVHYSMQSLTLD--VLGTSELLAKNIGNAGFNEIMECLPAEDS-----510
Db 297 SVQISKVRSRLTLDVKWVPSVMSYFQSGVNSVANSIWEELNPKSESSESRNVNDEQS 356
Qy 511 -----VKPNFGSDMNARKDITAKYIERRYARKKHADN--AAKLHSLCEAVKTRDIFG 561
Db 357 GVLASARPRPRDPIPIKERFINAKYVEKFKVQKLKVDGRGFSVTRQIWDVAVQNKVQL 416
Qy 562 LLQAY--AD-----GVDLTEKIPLAN-----580
Db 417 ALRLLITADANATTPEQVMGGTESWSSPLASAGALLRKNSLASQSGRRNRWVPSLL 476
Qy 581 -----GHEPDE-----TALHLAVRSVDRSTSLHIVDFLVQN 610

Db 477 SSPDPGSRGALSPVSRSPDAAGSGGIDKDLRGCSLLHVACQIGD---ISLIELLQY 533
Qy 611 SGNLDKOTGSGTALHVCCLTDNAECIKLLRGKASTEIANESGETPLDIA 661
Db 534 GAQINCVDLTGRPLHHCVLGNNSCAKLLJLTRGAKAGAVDKGKTPLECA 584
RESULT 10
US-09-270-767-42748
; Sequence 42748, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 42748
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42748
Query Match 6.6%; Score 347.5; DB 4; Length 225;
Best Local Similarity 41.0%; Pred. No. 5.1e-18;
Matches 87; Conservative 31; Mismatches 73; Indels 21; Gaps 6;
Qy 517 SDMNARKDITAKYIERRYARKKHADNAKLHSLCEAVKTRDIFGLLOAVADGVDLTEKI 576
Db 6 SSMEERYDFIRAKYAKRYVMRTCSDDNDLRCDLEQAVVADMSQLQVWAGADLTCCCL 65
Qy 577 PLANGHEPDETALHLAVRSVDRSTSLHIVDFLVQN-----SGNLDKOTGKGSTA 624
Db 66 P---SSDAGETALHLAVLREMGSTLHIVDFLQNMPPKGLNKATNPAGLLD-VTGG-NTA 120
Qy 625 LHYCCLDNACLLKLLRGKASIEIANESGTPDLDIKRLKHEHCELLTOALSGRFNSH 684
Db 121 LHLCALHRRRCMKLLLRSGADYELKNSQNTALDIKEMGHNCRLEICAIKREKSAF 180
Qy 685 VHVEYWRLLHED--LDSEDD--MDKLOPS 712
Db 181 DHINTDNLNPNEDGSTDFSDDETVIDERSSSS 212

RESULT 11
US-09-949-016-6235
; Sequence 6235, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6235
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6235

Query Match 5.3%; Score 278; DB 4; Length 802;

Best Local Similarity	20.1%;	Pred. No. 4.7e-12;
Matches	Conservative 147;	Mismatches 333;
	Indels	Gaps
QY	42	EALDVRMVLVYKMKSVKAINSSGLAHVENEQYTOALEKFG-----GNCVCRD 90
Db	20	ERLKCYPELERLTWKFKIDVIKDGVALISAMRNYSSAVQKFSQTLQSPQDFIGDTLTD 79
QY	91	DPDLGSAFLKFSVPTKELTALFKNLQIONMNNIISFPDLSLLKGDGLGVGDKLKPFDKAW 150
Db	80	EINIAESFKFAELNLNEVNERMMVMVFNASDLLIKPLENFRKEQI-GPTKRRKKKFKXDG 138
QY	151	KDVEYTKTK--TSKEKKEHAKLHGMIRTEISGAETIAEEMEKERRFFQLQWCEVLLKYNE 207
Db	139	ERYFSLDRHLHLSKKKE-----SQLEADL--QVDEKRNHPFESSLDVYVYQIE 187
QY	208	IKIKKGVDLLQNLKIYFHAQCNFFQDGL--KAVESLKPISIBTSLTDLHTIKQAODEERRQ 265
Db	188	VQSSKKNFIVEPVLAFHSL--FISNSLTVELTQDFLPYKQQLQLSLQTRNHFSSSTREE 245
QY	266	LIQRLDLKALQVEQKEDSQIROSTAYSLHQPOGNKEHGTER-NGSIY--KKSDGTRK 321
Db	246	MEELKKRMKEA-----POTCKLPQGTIEGVLYTOEKRWALGIS- 283
QY	322	VWQKRKCSXNGFGLTIS-----HGTANRPPAKNLLTC-QVKTNPEKK-CFDLISHDR 373
Db	284	-WKYICYQYKEYKTLTWTMPWEQKPGAKGGPDLTLKYCVARKTESIDKRCFQDIETNER 342
QY	374	--TYHQAQDEQECQIWMVSQNSKEBALNNAFKDDNTGNNIVQELTKETIISSEVQMT 431
Db	343	PGTITLQALSEANRRLWEAM-DGKEPIYHSPITKQEMELNEVGKPVKRCINIE-- 398
QY	432	GNDVCCDGGAPDTWLSTNIGILTICIECSGHIHRELGVHYSQMSQSLTDLVLTSELLAKN 491
Db	399	-----TKGIKT-----EGLYRTVGSNIQVKLL----- 421
QY	492	IGNAGFNEIMECCLPAESVKPNPGSDMNAKDYITAKYIERRYARKGHADNAAKLHSLC 551
Db	422	--NAFPDP--KC-----PGD-----VDFHNSD 439
QY	552	EAVKTRDIFGLQAYADGVDLTXEKIPLANGHEPDET-ALHILAVRSVDRSTLSHVDLVLQN 610
Db	440	.WDIKT-----ITSSLKFLRLNSEPVMTYRLHKELYSAK----- 474
QY	611	SGNLDQTKGKSTALH---YCLLTDNAECLKLLRGKASIEIANESGETPLDIAKRLKHE 667
Db	475	SDNLDYRLG---AIHSLVYKLPKKNREMLLELLIRHLNV-----CE 512
QY	668	HCRB-LLTQALSGRFNSHVHVEYEWRLHEDLDESDDMDKLOPSENRRDRPISFYQL 726
Db	513	HSKENLTP-----SNMGVIFPTLMRAQEDTVAAAMNLIKQ----- 549
QY	727	GSNQLQSNVSLARDAANLAKEKORAEPMPSIQNETYCALLSGSPPPAQAAPTSTTSAPP 786
Db	550	-----NIVVE-----ILIEHFGKIVLG--PPESA-----APP 575
QY	787	LPPRNV---GKVOTASS---ANTLWKTNSVSDGGSRQR-----SSSDP--PAWH 828
Db	576	VPPPRVTARHKPTITISKRLRRETRVFYTSLSDESEDEIQHTPENGITITSSIEPPKPPQH 635
QY	829	PPLPPLRVSTNPLTPTPPPVAKTPSPVMEALSQPSKPAP-----PGISQIRP 876
Db	636	PKLPIQRSGTDPGRKSPSRPI-----LDGKLEPCPEVDVGVKLVRLQDGGTKLTP 686
QY	877	P-----PLPPOPPSRRLPQ---KKPAPGDKSTPLTNKGQPRGVDLSATEALGPLSNAMVL 929
Db	687	KATNGMPWGSPTKTPSFHIKRPA-----RPLAH--HKEGDAD-SFSKVRPPCKEPII 738
QY	930	QPPA---PMPKRSQATKLPKRVKALYNCVADNPDELTFSS 966
Db	739	RPPVRPDPDCRAATPQKPEPKP---DIVAGNAGEITSS 774

RESULT 12

```

US-09-949-016-11166
; Sequence 11166, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11166
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11166

```

[illegible]

7.

DB 213 -----KAFMNNKFGIKIMLE-----TEEGI-----LLLVRA 238

THIS PAGE BLANK (USPTO)

Db 421 KEIISVQRMGTNDVCCGAPDPTWLTSTNLGILTCIECSGIRHGLVHYSRMQSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYIERRYARKKH 540
Qy 541 ADNAAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHALHLAVSVDRTS 600
Db 541 ADNAAKLHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHALHLAVSVDRTS 600
Qy 601 LHI VDFLVQNSGNDLKQTKGSTALHYCCLTDNAECLKLLRGRKASIEIANESGETPLDI 660
Db 601 LHI VDFLVQNSGNDLKQTKGSTALHYCCLTDNAECLKLLRGRKASIEIANESGETPLDI 660
Qy 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDWDEKLOPSENRREDRP 720
Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDWDEKLOPSENRREDRP 720
Qy 721 ISFYOLGNSLOSNVSLARDAANLAKEKORAFMPSILONETYGALLSGSPPPAQAAPS 780
Db 721 ISFYOLGNSLOSNVSLARDAANLAKEKORAFMPSILONETYGALLSGSPPPAQAAPS 780
Qy 781 TTSAPPLPPRVNKGVTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVTVSTN 840
Db 781 TTSAPPLPPRVNKGVTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVTVSTN 840
Qy 841 PLTTPPPPVAKTSPVMEALSOPKAPPGISQIRPPPLPPOPPSRPLPQKPAFGTKST 900
Db 841 PLTTPPPPVAKTSPVMEALSOPKAPPGISQIRPPPLPPOPPSRPLPQKPAFGTKST 900
Qy 901 PLTNKGQPRGVDLSATLGLPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCVADNP 960
Db 901 PLTNKGQPRGVDLSATLGLPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCVADNP 960
Qy 961 DELTFSEGDVIIVDGEEDQEWIGHIDGDPGRKGAFFVSFVHFAD 1006
Db 961 DELTFSEGDVIIVDGEEDQEWIGHIDGDPGRKGAFFVSFVHFAD 1006

RESULT 2
T42627
ADP-ribosylation factor-directed GTPase activating protein, isoform a - mouse
N:Alternate names: ASAP1a protein
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42627
R:Brown, M.T.; Andrade, J.; Radhakrishna, H.; Donaldson, J.G.; Cooper, J.A.; Randazzo, F.
Mol. Cell. Biol. 18, 7038-7051, 1998
A:Title: ASAP1, a phospholipid-dependent arf GTPase-activating protein that associates with
A:Reference number: 222178; MUID:99038209; PMID:9819391
A:Accession: T42627
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1147 <BRO>
A:Cross-references: UNIPROT:Q9QWY8; EMBL:AF075461; NID:G4063613; PID:G4063614; PIDN:AAC9
C:Genetics:
A:Gene: Shag1
C:Function:
A:Description: interacts directly with ADP-ribosylation factors (Arf1, Arf5 and to a lesser
A:Note: supposed to coordinate membrane remodeling events
C:Keywords: membrane trafficking

Query Match 56.9%; Score 3000.5; DB 2; Length 1147;
Best Local Similarity 53.9%; Pred. NO. 1.7e-138;
Matches 621; Conservative 128; Mismatches 232; Indels 171; Gaps 19;
Qy 1 MPDQISSEFVAETHEDYKAPTASSFTTTRTAQCRNTVAALFENLDVDRMVLKMKSKVKA 60
Db 21 MPDQISSEFVAETHEDYKAPTASSFTTTRTAQCRNTVAALFENLDVDRMVLKMKSKVKA 80
Qy 61 INSSGLAHVENEQYQTALEKFGGNCVCRDDPLGSAFLKFSVFTKELTALFKNLIQNN 120

Db 81 IYNSQDHVQNEENYAQVLDKFGSNFLSRNDPDLGTAFAVKFSTLTKELTSTLLKNLQGLS 140
Qy 121 NIISFPPLDLSLLKGLDKVKGDLKKPPDKANKDYETKTKIEKKEKHAUKLHGMRTEISG 180
Db 141 HNVIFTLDSLLKGLDKVKGDLKKPPDKANKDYETKTKIEKKEKHAUKLHGMRTEISG 200
Qy 181 AETAEMEKEKRRFQLOMCEYLLKKNVEIKIKKGVDLLQNLIKYPHACNFPQDGLKAVES 240
Db 201 AETAEMEKEKRRFQLOMCEYLLKKNVEIKIKKGVDLLQNLIKYPHACNFPQDGLKAVES 240
Qy 241 LKPSIETLSTLHTIKQAQDEERRQLIQLRDLKSLAQVBEKE 285
Db 261 LKQVIEKLAADLYNIKQDDEKKQLTALRDLIKSSQLQDPKEVGGLYVASRANSRRDS 320
Qy 286 QIROSTAYSILHOPGNKEHGTENGSLYKKSDDGIRKWKOKRCSVKNGFLTISHGTANRP 345
Db 321 QSRQG-GYSMHQLOGNKEYSGEKKGFLTKKSDGIRKWKOKRCAVKNKGLTISHATSNRQ 379
Qy 346 PAKLNLITCOVKTNPBEKKCFDLISHORTYHFOAEDEQECQIWMVSLQNSKEEALNNAFK 405
Db 380 PAKLNLITCOVKNPAEDKKSFDLISHNRTYHFOAEDEQDYIAWISVLTNSKEEALTWAFR 439
Qy 406 GDNNTGNNIVQELTKIEIISVQRMGTNDVCCGAPDPTWLTSTNLGILTCIECSGIHRE 465
Db 440 GEQSTGENSL-EDLTKAIEEDVQRLPGNDICCDGSGSEPTWLTSTNLGILTCIECSGIHRE 498
Qy 466 LGVHYSQMSITLDTVLGTSSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDY 525
Db 499 MGVIHSIQSLDLEDKLGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDY 557
Qy 526 ITAKYIERRYARKKHADNAALHSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPD 585
Db 558 ITAKYVDRHRSKTCASSAKNELLEAIKSRDLLALIOYVAGVELMEPL-LFPGQELG 616
Qy 586 ETALHLAVRSVDRTSLHIVDFLVQNSGNDLKQTKGSTALHYCCLTDNAECLKLLRGA 645
Db 617 ETALHLAVRTADQTSHLVDFLVQNSGNDLKQTKGSTALHYCCLTDNAECLKLLRGA 676
Qy 646 SIETANESGETPLDIKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDW 705
Db 677 TVDIVNGETALDIKRLKATQCEDELLSQAKSGKFNPHVHVEYEWRLHEDLDESDDW 736
Qy 706 DEKLQPSENRREDRPIISFYOLGNSLOSNVSLARDAANLAKEKORAFMPSILQNETYGA 765
Db 737 DDKEPPIKGRSPRQSFCHSSSISPD--KLAALPGFSTPRDKQL------SVGA 784
Qy 766 L---LSGSPPPAQAAPSTTSAPPLPRNVGK-----VQTASSANTLWKTNSV 810
Db 785 FTNQIFASTSTDLPTSP-TSEAPPLPRNAGKGTGPPSTLPLGTQTSSGSTLSKKRPP 843
Qy 811 SVDGSGRQSSSDPPAVHPPPLP----- 833
Db 844 PPPFG-HKRTLSDPPSPPLPHGPPNKGAI PMGNDVGPSSSKTANKFGLSQOASTSSAKT 902
Qy 834 -----LRVTSTN----- 850
Db 903 ALGRPVRLPKLPQKVALRKTTSHLSLDRTNI PPETFQKSSQLTELQKPPGLGELPKPV 962
Qy 851 --AKTPSVMEALSOPSK-----PAPPGISQIRPPP-----LPP----- 881
Db 963 ELAPKPVGELPPKPGELPPKPGQLGDLPPKPOLSDLPPKQMDLPPKPKQLGDLAKSQA 1022
Qy 882 -----QPPSRLPQKPAFGTKSTPLTNKGQPRGPDLSATLGLPLSNAMVLQPPAP 934
Db 1023 GDVSAKVQPPSEVQTR-----SHTGLSPNVQSRDAIQKQASDESDNLTFTLP-ETVVP 1075
Qy 935 MPRKSQATKLKPKVKALYNCVADNPBELTFSEGDVIIVDGEEDQEWIGHIDGDPGRK 994
Db 1076 LPRKINTGKVRVKTIYDQADNDDELTFIESEVIIVTGEEDQEWIGHIEGQPERKG 1135
Qy 995 APFVSFVHFAD 1006
Db 1136 VFPVSFVHILSD 1147

RESULT 3

A96634
probable GCN4-complementing protein F23C21.2 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96634
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A96141; MUID:21016719; PMID:11130712
A:Accession: A96634
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <STO>
A:Cross-references: UNIPROT:Q9C6C3; GB:AE005173; NID:g10092284; PIDN:AAG12697.1; GSPDB:C
C:Genetics:
A:Gene: F23C21.2
A:Map position: 1

Query Match 9.2%; Score: 485.5; DB 2; Length 776;
Best Local Similarity 24.4%; Pred. No. 2.8e-16;
Matches 185; Conservative 131; Mismatches 284; Indels 157; Gaps 25;
Qy 51 LYK-MKSVKAIN--SSGLAHVENEEOYTOALEKFGNCVCRDDPD--GSAFLKFSV 103
Db 34 LYGVKFKMGALGEASTGV-----APADSLEEFAG--HDDPVSVISGPGVSKFIN 84
Qy 104 FTKELTALFKNIQNMNNIISPLDLSLKGDLKGVKGLKPPDKAWDKDYETKTKIEKE 163
Db 85 TLRELSYKFEKLSQVEHVLRLTNFTVDLQEAQ--ESRRFRDVAHSHYDQAREKPVSL 143
Qy 164 KKEHAKLHGMTRTETSGAIEAEMEKEKRRFQLOMCEVLLKVKNEIKKKGVLDLQNLKY 223
Db 144 KKN-----TRGDIV--AELEDLNSKSAFKSRFNLSNMTAEAKKYEFLESISAI 195
Qy 224 FQAQCNFFQDLKAVESLKPSETL-----STDLTHTIKQAQDEERRQLIQLRDLKSL 277
Db 196 MDSHPKYFKGLVDLSQLEPIYHVLTYAQQSKESKIQDFPAQRIQEFRTQSLDSQ 255
Qy 278 QVEQKEDSQIROSTAYSILHQGNKEHGT-----BRNGSLYKSGDKIRKVMQKKS 329
Db 256 ASAKADPSDVGNGHVYRA--IPKKNVEANSVSTADKEVTQGYLLKRSASLRADWKRPFV 314
Qy 330 VKN-CFLTISHGTANRPAK-----LNLT 353
Db 315 LDNHGSLYYRNTGNKSAKSHYSGLEHSGVGFGRFRTRHNRASQSGSLDCNNMIDLT 374
Qy 354 QVQKNPBE---KKCFDLISHDRTHFOAEDEQECIWMVSLQNSKEBALNAPKGGD-- 408
Db 375 SLIKLDAEDTLRLCFRIISQKTYTLQANGADRMVWNTAITAILNSHFLOQSPA 434
Qy 409 -----NTGENNIVQELT-----KEIISVQRTMGNDVCCDCGAPPTWLS 448
Db 435 RYLDKNTSSGPATENLTNQKEDYNQRLNVDGDDVLTILRETPGNNTCAECNAPPDWAS 494
Qy 449 TNLGTLTCEGSHIRELGVHVSQMSITLD--VLGTSSELLAKNIAGNFEIMECCLP 506
Db 495 LNLGVMCEGSGVHRNLGVHVSRTLDLVKWEPTILDRLNLGNGYCNVWEELH 554
Qy 507 AEDS-----VKPNPGSDMNARKDYITAKYIERRYARKKHAD--NAAKLHSLCE 552
Db 555 HLDDSEKSGTDTLASVSKPSSDFTLKEKYNGKYLEKALVVKDREANSTASSRIWE 614
Qy 553 AVKT---RDIFGLL-QAYADGV-----DLTEKIPLANGH--EPDE----- 586

Db 615 AVQSRNRIYRLIVKADANIINTKFDITDLVDVYHHHVDAPDEVKKRHDNDACQRIKN 674
Qy 587 -----TALHLAVRSVDRTSLSHIVDFLVQNSGLNDKQTKRGSTALHYCCLTIDNAE 635
Db 675 SNEARNCLQCGSLLHVAQSGDPI--LLELLQFGADINMRDYGRTPLHHCIAAGNNA 731
Qy 636 CLKULLRKGASIEIETANESGETPDLQAKRLKHEHCEEL 672
Db 732 FAKVLLRRGARPSIEDGGGLSVLERAMEMGAITDEEL 768
RESULT 4
T26508
hypoetical protein Y17G7B.15 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26508
R:Smyle, R.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z20225
A:Accession: T26508
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-752 <WIL>
A:Cross-references: UNIPROT:Q9XXH9; EMBL:AL023828; PIDN:CAA19462.1; GSPDB:GN00020; CESP
A:Experimental source: clone Y17G7B
C:Genetics:
A:Gene: CESP:Y17G7B.15
A:Map position: 2
A:Introns: 177/3; 221/3; 345/1; 539/1; 654/2; 691/2; 740/1

Query Match 9.1%; Score 479.5; DB 2; Length 752;
Best Local Similarity 25.7%; Pred. No. 5.3e-16;
Matches 178; Conservative 92; Mismatches 206; Indels 217; Gaps 26;
Qy 144 KPPKAWKDYETKTKIEKEKHAHGMTRTETSGAE-----IAEMEKERRFPFQIQ 197
Db 40 EPFSRSTKSGIRTSKV-----ITRTTISAEITKLMDLLEQORREKR-VKLE 86
Qy 198 M-----CEYLLKVKNEIKK-KGVLDLQNLKYFQAQCNFFQDLKAVESLKPSETL-- 248
Db 87 IEKSRIRAEINQRMQEIISSEPTKRGSKSLMR-----RAAKTKGRISLLI 133
Qy 249 ---STDLTHTI-KQAQDEERRQLIQLRDLKSLAQVEQKEDSQIROSTAYS-----LHQPQ 299
Db 134 RDSVDSGSIITDKALKOKKRG--SLEWEVKDQGRFAESFSDSLMWMFQHP 184
Qy 300 QNK-EHGTFRNGSLYKSGDKIRKVMQKKSQNGKGLTISHGTANRPAKL---NLLTQ 355
Db 185 GMPTEPDVWMEGYLYKRSNAPKTNRRWFQIKQKQLLYSHRSTOLEPATIMEENLRICL 244
Qy 356 VK---TNPEKKCFDLISHDRTHFOAEDEQECIWMVSLQ-----NSK 396
Db 245 VRPAPSNIDRTGCFELVPTTRIHLQADSESLCDMMRALQRTILALHEGSDVAVSTSP 304
Qy 397 EEALNAPKGGDNTGENNI-----VQELTK-----E 422
Db 305 RNKTTSSSGVLTLSANAIPLSNAMDVTKGRSVSDPASTYTSANTSSISTAAGSSSTT 364
Qy 423 IISVQRTMGNDVCCDCGAPPTWLSNLGTLTCEGSHIRELGVHVSQMSITLDVLG 482
Db 365 AFEQVRVPNGEVNACDCGSPAPKWSINLGVVLTECSGAHRSIGVQTSKVRSLCMTSID 424
Qy 483 TS--ELLIAKNIAGNFEIMECCLPADSDSKVP---NPGSDMNARKDYITAKYIERRY- 536
Db 425 NELRDVLLA--LGNRQVNEIFLAHLPADSIIVPPQINEKSARPAAREAWIKAKYVERRFAV 482
Qy 537 -----RKCH-----AD-- 542
Db 483 AEDTRASSATNRQEHKHKHTSIGNSSNGVNRSSSYADVQDAESGGLLDADPHSADLS 542
Qy 543 ----NAAKLHSLC-----EAVKTRDIFGLQAYADGVDLT 573

```
Db 543 VPVTSKRLSACGSDTNLDALGSSIDTKTVEWDSVKEACECGDLLALMTAYAGFDLN 602
Qy 574 EKIPLANGHEPDETHALHVAHSVDRSTLHIVDFLVQNSGNLDKQGTGSTALHYCCCLTDN 633
Db 603 ---ALHNG---TTALHTATRNQGTAA---VEFLLNGAKINMLDEKLTPLHLAAKEGH 652
Qy 634 AECLKLLRGKASIEIANESGETPLDIAKRLKH 666
Db 653 TLPVCQLLRGAGNSLANVDSKTPLDIAWECTH 685

RESULT 5
T48577
hypothetical protein T31B5.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T48577
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224490
A:Accession: T48577
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-750 <BEV>
A:Cross-references: UNIPROT:Q9LYU6; EMBL:AL163491
A:Experimental source: cultivar Columbia; BAC clone T31B5
C:Genetics:
A:Map position: 5
A:Introns: 23/2; 57/1; 80/3; 101/3; 123/3; 139/3; 171/3; 185/3; 257/3; 334/3; 363/3; 475
A:Note: T31B5.120

Query Match 8.8%; Score 466; DB 2; Length 750;
Best Local Similarity 22.8%; Pred. No. 2.4e-15;
Matches 180; Conservative 122; Mismatches 302; Indels 186; Gaps 26;

Qy 40 IEEALDVRMVLVYKMKSVKAINSSGLAHV-ENEEQYQALEKFGNGVCVRDDPDL---- 94
Db 1 MEESAIEIURSLKPYKCRKY-TEGLGEAYGDIAFASALETFGGG---HNDPISVAFG 56
Qy 95 GSALFVSVFTKELTALPKNLTQNNNIISFPLDSLKGLDGKVGKDLKPKFPDKAWKDYE 154
Db 57 GPVMTKFTIALREIGTYEVLRSQVEHILNDRLLQFANMDLHEVKA-----R 103
Qy 155 TKITKIEKEKHAHLGHWIRTEISGAIEAEMEKERRFPQLQMCYELLKKNELKIKGV 214
Db 104 EKFLSLRKGTKSDV-----AAALEOELHTSRMFEQARENLVTALSNVEAKKRF 152
Qy 215 DLLQNLIKYFHAQC�FFODGLKAVESLKPSITETSLDLHTTIKQAOBERRQLIQLRDLK 274
Db 153 EFLEAVSGTMDAHLRYFKOGVELLHQMPEYINQVLT-----YAQSRERSNYEQALNE 206
Qy 275 SALQVEQKEDSIQSTAYSILHQPQCNKEHGHTERNQ-SLYK-----KSDGIRK 321
Db 207 KMQEYKRVQDRSRWGSNGSPNGD---GIQAIGRSHKWDIAVMQSAARGKSGSQL 263
Qy 322 VWQKRCVKGNGFL-----TISHGTANRPAPK-----LNLTCQVKNPPEE---KKCFD 367
Db 264 SGQRNSSELGSLLRWLSNNHGHGCVHDEKSVARHTVNLITSTIKVDADQDLRFCFR 323
Qy 368 LISHDRTYHFOADEQEQIWM-----SVLQNS-KEEALNNAFKGDDNTGENNIVQE 418
Db 324 IISPTKNTLOAESALDQMDWIEKITGVIASLSQVPEQRLPGSPMG---SGHRSASE 380
Qy 419 -----LTKEIISE-----VQRMGTNDVCC 437
Db 381 SSSYESSEVDHPTTEEFVCSFLGYNRPSPRSPQRSIRKGEKPIDALRKVCNDKCA 440
Qy 438 DCGAPDPWLSNLGILTCIECSGTHRELGVHVSQMSLTLD--VLGTSELLAKNIGNA 495
Db 441 DCGAPEPDWASLNLGVLCIECSGTHRELGVHVSQMSLTLD--VLGTSELLAKNIGNA 500
Qy 496 GFNEIMECCLPADSVKPNPG-----SDM-NARKDIYITAKYIERRYAR 537
```

```
Db 501 FANTVWEELLHRSATIHFDPGLTVSDKSRVMTGKPSYADMISIKYIQAKYAEKLFVR 560
Qy 538 KKHADN--AAKLHSLCEAVKTRD---IFGLL-----QAYADGVDLTEKIPLAN 580
Db 561 RSRSDPFPQAAQQMDAVSGNDKKAVYRLIVNGDADVYVYDQTSSSLTSRLVILPE 620
Qy 581 GHEPDETHALHVAHSVDRSTLHIVDFLVQNSGNLDK 616
Db 621 RPKREDVLLRLRNELLORTGSSSNISPEGSGSLLHCACEKADLQWVLLQYGANVNA 680
Qy 617 QTGKGSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKRLKHCHCBELLTQA 676
Db 681 SDSSGQTPHCCLLRGKVTIARULLTRGADPEANNRREGKTALDIAESNPTDPEVL--A 737
Qy 677 LSGRFNSHVH 686
Db 738 LLSDTNGYNH 747

RESULT 6
D86242
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D86242
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroo, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86242
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-531 <STO>
A:Cross-references: UNIPROT:O04097; GB:A8005172; NID:g1931654; PIDN:AAB65489.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 7.6%; Score 399; DB 2; Length 531;
Best Local Similarity 24.2%; Pred. No. 2.8e-12;
Matches 137; Conservative 85; Mismatches 193; Indels 152; Gaps 15;

Qy 212 KGVLLQNLIKYFHAQC�FFODGLKAVESLKPSITETSLDLHTTIKQAOBERRQLIQLRD 271
Db 3 QGYDLLNQLEPIHILTYAQO-----SKEQSKIEQDLARIQIEFRITQS 47
Qy 272 ILKSALQVEQKEDSQIR-----QSTAYSL-----HQPQCNKEHGHT----- 306
Db 48 ELDSQQLVANAESSGANGNRVCGNIPYKNTETSLTADKEGSHHHVSGSDHNTGVGFRFR 107
Qy 307 ---ERNGLYKSGGIRKWKQKRCVKNQGFLLTISHGTANRPAPKLNLLTCQVKNPPE- 362
Db 108 ARHNRSGSLTSGSLGYNTI-----DURTSILKLDADM 140
Qy 363 --KKCFDLISHDRTYHFOADEQEQIWMVSVLQNSKEBALN----- 401
Db 141 DLRLCFRIISPKQYTTLOAENGADRMMDWNKITKAITGLLNSHFLQQSPRYLDKDNSS 200
Qy 402 ---NAFKGDDNTGNNIVQELTKELISEVQRMGTNDVCCGAPDPWLSNLGILTCIE 458
Db 201 APANAVVSGDQIRHNSRFQNIQDDVSTILRGLPGNACAECAEPDPAWASLNLVLLCIQ 260
Qy 459 CSGTHRELGVHVSQMSLTLD--VLGTSELLAKNIGNAGFNEIMECCLPADSV----- 511
Db 261 CSGVHRLNGLVHISKVRSLSLDVKVWEPTILDFRNLGNVYCNLSWEGLHLLDDDCEDGSA 320
```

QY 512 -----KPNPGDMNARKDYITAKYIERARYARKKHAD-NAAKHLHSLCEAVKT---RDIF 560
DB 321 LSHASVSKPCBEDSPFVKEKYLILGKYLEKALVYKDESEANLSAASGRINEAVOSNRIREY 380
QY 561 GLUQAYAD-----GVDLTKIPLANGHEPDE----- 586
DB 381 RLIVITGDVNIINTKFDITDIDAYHHIDAEEK-AVKKEHDITVCORIKESNEPRSCLOG 439
QY 587 -TALHLAVRSVDRTHLHVDFIVQNSGNLDKGTGKSTALHYCCITDNNACLUKLLLRGA 645
DB 440 CSLHLVACHIGDSV---LLELLQFGADLNIRDYHGRTPLRHCISGNHKKFAKILLRGA 496
QY 646 STEIANESGETPLDIATKRLKHEHCEEL 672
DB 497 RPSIEDDGLSVLERAMENGAITDEEL 523

RESULT 7

T49496
hypothetical protein B14D6.480 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49496
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49496
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1087 <SCH>
A:Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.480
A:Experimental source: BAC clone B14D6; strain OR74A
C:Genetics:
A:Gene: NCSP:B14D6.480
A:Map position: 6
A:Introns: 67/3; 99/3; 270/2; 589/2; 616/2

Query Match 7.5%; Score 393.5; DB 2; Length 1087;
Best Local Similarity 21.7%; Pred. No. 1.3e-11;
Matches 148; Conservative 119; Mismatches 233; Indels 181; Gaps 24;
QY 34 RNTVAIEEALDVRMVLVYKMKSKVKAI-NSSGLAHVENEEOYTOALEKPE-----GGNCV 87
DB 334 RATMKALEQKQTNMRM---QMKRLIKRAEQVYAAQTEANDAFLOFMEALRDVYSTNANAV 390
QY 88 CRDDPOLGSAFLKFSVFTKELTALFKNLQNMNIIISFPLDLSLLKGLDKGVKGLKCPFD 147
DB 391 ---QPAIEHYFKIA---WEILRPERQNTQNLQKIVIDPMNKLYQLDIIKQAEAK-KRDPF 443
QY 148 KAWKDYEYTKITKIEKEKEHAKLHGMIRTEISGABIAEBEMEKERRFFQLOMCCEYLLKVN 207
DB 444 BESKDFYAVSVYLGQR-----HDSVKAKQSDS---KYQTKRKNFELKRPDYSSFMQD 493
QY 208 IK-IKGVVDLLQNLKYFHAQCNFPQDGLKAVESLKPSIETLSTDLHTIKQADEEROL 266
DB 494 LSGGRKEQIEILSHLTAKYADAQAKAFLTTAKYVEGLLPQLEALSTEVOEADKTYQYQRRR 553
QY 267 IOLRDLKALQVEKEDSIQSTAYSILHQPQGNK-----BHGTENRSLVK--- 314
DB 554 EKKRLLLE-----KGNLAYAPEPEPSTSSGSPSTGQNGTAKGELLWALNRP 600
QY 315 -----KSDGIRKVMQKRCVKNGFLTISHGTANR-----PPAKNLNLT 353
DB 601 GGHVDPRLANKQGWKFW-----IVLDQKLSSEYNNWKQRLDLHMDPIDLRAS 649
QY 354 CQVKNPBEKKFDLISHDRTHVFOADEQECQIWMVSLQNSKEAL----- 400
DB 650 VREARNAERRFCFEVITPQKRYQATSEEDWNWIMAINNALQSAVEGRGLKMKPPPS 709
QY 401 -----NNAFKGDDNTGENNIVQELTK----- 421
DB 710 SIGHSDFRDRDIASVLTKGSSINHHHHFHGHNSQNLSPISRRRTTVGARPSTARAPSF 769

QY 422 -----EIISEVORWITGNVCCDCCGAPDPT-WLSTNLGILTCIESGIIHRELGVHYSF 472
DB 770 DESPDRLLOMLRENDQ--GNSWCADCGSSNKVWVSLNLAITVCIESGIIHSLGTHISK 827
QY 473 MOSLTDLVLGTS-----ELLAKNIGNAGFNEIMECCLPAEDSVKPNPGSDMMARKDYITA 528
DB 828 VRSULTDITSTFTPDIIELMLL--VGNRVSNVWYEARLPDQKLVFQASREQRLR--FITA 883
QY 529 KYIERY-----ARKKHADNNAKHLHSLCEAVKTRDIFGLLQAYADGVDLTEKIPLAN 580
DB 884 KYVDRAFVEPISSTLSRYPTAE---ETLLAAIKKEIQVIVYALA-----LKAN 929
QY 581 CHEPDET-----ALHLAVRSVD 597
DB 930 PNITDKSRGTHAVYLALAAAD.950

RESULT 8

T39731
cbx2 protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jul-2004
C:Accession: T39731
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21874
A:Accession: T39731
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-870 <WOO>
A:Cross-references: UNIPROT:Q9UUE2; EMBL:AL109846; PIDN:CAB52806.1; GSPDB:GN00067; SPDB:
A:Experimental source: strain 972h-; cosmid c17G9
C:Genetics:
A:Gene: SPDB:SPBC17G9.08c
A:Map position: 2
A:Introns: 13/1; 57/3; 118/3; 166/3; 709/1

Query Match 6.4%; Score 335; DB 2; Length 870;
Best Local Similarity 22.2%; Pred. No. 7e-09;
Matches 147; Conservative 111; Mismatches 254; Indels 150; Gaps 27;
QY 34 RNTVAIEEALDVRMVLVYKMKSKVKAI-NSSGLAHVENEEOYTOALEKFGNCVCRDDP 92
DB 256 RATILNYERTTHDMRMVLKCLIKRIEVAHSHGLLYMSYKE-LMSAFER-----VATINP 309
QY 93 DLGSAPL-----KFSVFTKELTALFKNLQNMNIIISFPLDLSLLKGLDKGVKGLDK 143
DB 310 PAFKFLDHYAAQAQSFDSFNIDRARLLRFLIE-----PLRKIYDTDINKV-STTK 361
QY 144 KPPDKAWDYETKITK-IEKEKEHAKLHGMIRTEISGABIAEBEMEKERRF-----FOL 196
DB 362 KDFBETSRDYVTSLSRYLSKSEK-----ETSSDKTKESKFAAKKRDPEL 405
QY 197 QMCVEYLLKVNIEK-IKGVVDLLQNLKYFHAQCNFPQDGLKAVESLKPSIETLSTDLHT- 254
DB 406 SRFDYYSYMODINGRKGQEVLSVLTSPAANDYNIHSLTDIDALRPSIIQLQ-DIVTE 464
QY 255 -----IKQAODEERRQLIQLRDLILKLSALOVEKEDSOIROSTAYSILHQPQGNKEHGT 308
DB 465 ANKEPQLRAEERREERYIE-----TSRELEDK-DAEI-AAQAYKAVDQDT-----SAK 513
QY 309 NGSL--YKSKDGIRKV-----WQREKCSVQNGFLTISHG-----TANRPPAKL 349
DB 514 QGLLAFLLSKSTDLQVVGKSGWHKY-----WVLDHGKICEYANNKQSLHTEPIDL 566
QY 350 NLLTCQVKNPBEKKCFDLISHDRTHVFOADEQECQIWMVSLQNSKEALANNAFKGDD- 408
DB 567 LMATVRPAQSVSRKFCFEVITPQTKRYQATSKAEMHSHWIEAIOYSISESIQVKGKGTSM 626
QY 409 NTGENNIVQELTKELTISVQRM-----TGNDVC 436
DB 627 NSEETSVKHGPTSTIGKALQRVASVTSRSHNSDSKEKKQTKSPSLVLTLEMHSSDQSC 686

Qy	612	GNLDKQTKGSGTALHYCCLTONTAECLKLLRGKASIEITANSGETPLDIAKRLK-HEHCE	670
Db	822	-----TKDGRTVLHLAASIGSVELAQQLIWHNADAQILDNNRGRCLFYARNGRFVFD	875
Qy	671	ELLTCAALSGRNSHVHVEYWRLLHEDIDESDDMDDEKLQPSNNRDRPISFYOLGSNQ	730
Db	876	MLVTVAGLSPDYQ-----LPQEIINFDSQMP-----FFAMGSTS	908
Qy	731	LQSNVSLIARDAANLAKEKQAFMPSILQNETYGA-LLSGSPPPPAQAPSTTSAPPLPP	789
Db	909	NRDYSTS-----GDEYSGRRISMGPFPQVPARRYLPQPOPELDE	947
Qy	790	RNV 792	
Db	948	TSV 950	

RESULT 11

T26737
hypothetical protein Y39A1A.15a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26737
R:Wall, M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z20257
A:Accession: T26737
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-975 <WIL>
A:Cross-references: UNIPROT:Q9XX14; EMBL:AL031633; PIDN:CAA21026.1; GSPDB:GN000021; CESP:
A:Experimental source: clone Y39A1A
C:Genetics:
A:Gene: CESP:Y39A1A.15a
A:Map position: 3
A:Introns: 105/3; 135/1; 211/3; 257/1; 454/3; 496/2; 870/3; 924/3

Query Match	6.1%;	Score 319.5;	DB 2;	Length 975;
Best Local Similarity	21.6%;	Pred. No. 4.6e-08;		
Matches	143;	Conservative	83;	Mismatches 226;
			Indels	211;
			Gaps	21;
Qy	232	QDGLKAVESLKPSTETLTDLHTIKQADERRRLOLIQRLDLKSALQVQKEQSQIRQST	291	
Db	421	QSGANQVSASTSHLPTSPNT--QRKNRRISNIFPKD-----HOEKSQMIESL	470	
Qy	292	AYSLLHQPQGNKEHCHTERNGSLYKKS--DGRKVKWQKR-KCSVKNGFLTI-----	337	
Db	471	NLGIGRAIPKQ-----GNDYKKSXSALNREWKYVCLYSDGRUTYTNLKEYMDKT	524	
Qy	338	-----SHGTAN-----	343	
Db	525	AHGKEMDKLATIRITGRHLPHLSHRVASSVDPMNNGGGGGTPTLKSYEPRRSDVGANS	584	
Qy	344	-----RPPA-----KLNLTTCQVKTNPBEKKCFDLISHD	372	
Db	585	DGTSGGSDDAIKENQRQHFSPAMPQTQVAGKKRESRKIGTNSKHNDDECEFEVINNC	644	
Qy	373	-RTHFOAEDEQECQIWMVQLQNSKEBALNNAFKGDDNTGNNIVQLTEIISVEQVMT	431	
Db	645	LMWFEFCAGSLEERDEWIAIGGIEIKSLG---KEVANAKTNRA-VANRPDIAALRSIP	700	
Qy	432	GNDVCCDCGADPTWLGTNLGILTCIECSGTHRELGVHYPMSQSLTLDVLGTSELIAKN	491	
Db	701	GNGRCADCGNPSSWASINIGIICIECSGIHRNLGSHISKVRGLELDQWPVEHLAVMQA	760	
Qy	492	IGNAGFNEIMECCLPABSDSVKPNPSSDMNARKDYITAKYIERRYARKKHADNAAKLHSLC	551	
Db	761	IGNDKANEMWEFGL-LNGERKPTPSSREKERFIDRKYVQKAPL-KPIASGEVPTSQLI	818	
Qy	552	EAVKTRDIFGLQLAYAGVDLTETKPIPLANGHEPDETALHLAVRSVRTSLHIIVDFLVQNS	611	
Db	819	SAVLARDVMSL-----NVLNANG-----MSVEINTT-----	845	

RESIST 12

T38398
hypoetical GTPase activating protein for Arf protein - fission yeast (Schizosaccharom
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T38398
R/McLean, J.; Harris, D.; Bartell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A/Reference number: Z21791
A/Accession: T38398
A/Status: preliminary; translated from GB/EMBL/DBDJ
A/Molecule type: DNA
A/Residues: 1-923 <MCL>
A/Cross-references: UNIPROT:Q10165, EMBL:Z69240; PFDN.CAA93233.1; GSPDB:GN000066; SPDB:S
A/Experimental source: strain 972h-; cosmid 26A3
C/Genetics:
A/Gene: SPDB:SPAC26A3.10
A/Map position: 1
A/Introns: 6/1: 47/3

[illegible]

Qy 390 -----SVLQNSK-----EEALNNAF--KG-----D 407
 Db 636 FQLSRVATPLASVVRPSKVPPLFSTSPETTPISRLSGSGIKAFPSRKGSGWNLOQFPRS 695
 Qy 408 DNTGNNIVQ----ELTKIIESEVQMT--GNDVCCDCGA-PDPTWLSTNLGILTCIBCS 460
 Db 696 DMSGTHMEQLERYHASANIFIQMLRKTDVNSVCADCGSKVDVTWCSINIPVVLICIBCS 755
 Qy 461 GIHRELGVHSPMQSLTLDVLTGSELLAKIGNAGFNEIMECCLPADSDVKPNPGSDMN 520
 Db 756 GIIHSLGTHISKTRSLDLSLQSKVLLCKIGNAANVRVYKGL--SNPSLKPKPEHNAQ 814
 Qy 521 ARKDYITAKYIERRYARKKHADNAKLSLCEAVKTRDIFGLLOQAYAGVDLTTEKIPLAN 580
 Db 815 VKLAFQAKKYVEHAFIDFAGVDADA---TLGLEQNKISKILLGLAAKPNFEE-----N 866
 Qy 581 GHEPDETALHLAVRSVDRSTSLHVDLQVN 610
 Db 867 G-----VVFLKAVTRDTSKLHLELFPWN 890

RESULT 13
 H59434
 C:Species: Homo sapiens (man)
 C:Date: 03-Jun-2002 #sequence revision 03-Jun-2002 #text_change 09-Jul-2004
 C:Accession: H59434; A59435; B59435; C59435
 R:Billuurt, T.; Der-Sarkissian, H.; Billuurt, P.; Tissot, M.; des Portes, V.; Bruls, T.;
 Eur. J. Hum. Genet. 5, 105-109, 1997
 A:Title: Mapping of the X-breakpoint involved in a balanced X;12 translocation in a female
 A:Reference number: H59434
 A:Accession: H59434
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-802 <B1>
 A:Cross-references: UNIPROT:060890; GB:NP_002538; PID:g4505507; PIDN:NP_002538.1
 R:Billuurt, P.; Bienvenu, T.; Ronce, N.; des Portes, V.; Vinet, M.C.; Zemmi, R.; Roest C
 A.; Morsine, C.; Chelly, J.
 Nature 392, 923-926, 1998
 A:Title: Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation
 A:Reference number: A59435
 A:Accession: A59435
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-802 <B1>
 A:Cross-references: GB:NP_002538; PID:g4505507; PIDN:NP_002538.1
 R:Tentler, D.; Gustavsson, P.; Leisti, J.; Schueller, M.; Chelly, J.; Timonen, E.; Amner
 Eur. J. Hum. Genet. 7, 541-548, 1999
 A:Title: Deletion including the oligophrenin-1 gene associated with enlarged cerebral ve
 A:Reference number: B59435
 A:Accession: B59435
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-802 <TEN>
 A:Cross-references: GB:NP_002538; PID:g4505507; PIDN:NP_002538.1
 R:Billuurt, P.; Chelly, J.; Carrie, A.; Vinet, M.; Couvert, P.; McDonnell, N.; Zemmi, R.;
 Ann. Genet. 43, 5-9, 2000
 A:Title: Determination of the gene structure of human oligophrenin-1 and identification
 n.
 A:Reference number: C59435
 A:Accession: C59435
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-802 <B12>
 A:Cross-references: GB:NP_002538; PID:g4505507; PIDN:NP_002538.1

Query Match 5.3%; Score 278; DB 2; Length 802;
 Best Local Similarity 20.1%; Pred. No. 3.7e-06;
 Matches 201; Conservative 147; Mismatches 333; Indels 318; Gaps 47;

Qy 42 EALDVRMVLVYMKKSVKRAINSGLAHVENEQYTQALEKFG-----GNCVCRD 90
 Db 20 ERLKVEQELERTNFKIDVKDGNALISAMENYSSAVKFSQTQLSQFQDFDGTDLTD 79

Qy 91 DPDLGSAPLKFVSFTKEITALTAKNLIONMNNIIISPLDLSLLKGLKGVKGLKKPFDKAW 150
 Db 80 EINTAESFKFAELLNEVENEMMMVHNASDLLTKPLENFKEQI--GFTKRRKKKFKDQ 138
 Qy 151 KDYETKITK----IEKKEHAHLGMRITETISGAIABEEMEKERRFFQLOLCEYLLKVNE 207
 Db 139 ERFYSLLDRLHLHLSSKKKE-----SQLQEADL--QVDKERHNPFESSLDDVYQIQE 187
 Qy 208 IKIKKGVDDLQNLKYFHAQCNPFDQGL--KAVESLKPSTETLTDLHTIQAQDEERQ 265
 Db 188 VQESKKFNIVEPVLAFHLSL--FISNLSLTVELTDQFLPYKQQLQLSLQNTNHFSTREE 245
 Qy 266 LIQLRDILKALQVEKEDSQIROSTAYSLHQPGNKEHGTER--NGSLY---KKSDGIRK 321
 Db 246 MEELKKRMKEA-----POTCKLPQOPTIEGYLYTOEKWALGIS- 283
 Qy 322 VWQRRKGSVKNGLFTIS-----HGTANRPPAKNLLTC--QVKTNPBEKK--CFDLISHDR 373
 Db 284 -WVKYCYQYKEKETKLTMTWMEQKPGAKQGPLDLTKYCVRRKTESIDKRCFCFDIETNER 342
 Qy 374 --TYHFOAEBDEQCIIWMSVLQNSKEALNNAFKGDDNTGNNIVQELTKIEIIEVQRM 431
 Db 343 PGTTITLQALSSEANRLWMEAM--DGKEPTIYHSPITKQOEMELNEVGFVKRCINIE--- 398
 Qy 432 GNDVCCDCGAPDPTWLSTNLGILTCIBCSGTHRELGVHSPMQSLTLDVLTGSELLAKN 491
 Db 399 -----TKGIKT-----EGLYRTVGSNIQVQKLL----- 421
 Qy 492 IGNAGFNEIMECCLPADSDVKPNPGSDMNARKDYITAKYIERRYARKKHADNAKLSL 551
 Db 422 --NAFFDP--KC-----PGD-----VDFHNSD 439
 Qy 552 EAVKTRDIFGLLOQAYAGVDLTTEKIPLANGHEPDET--ALHLAVRSVDRSTSLHIVDFLQVN 610
 Db 440 WDIKT-----ITSSLFYLRNLSEPTWYRLHKLKELVSAK----- 474
 Qy 611 SGNLDKQTKGSTALH--YCCLTDNAECLKLLRGRASIEIANESGETPLDIAKRLKHE 667
 Db 475 SDNLDYRLG----AIHSLVYKLPEKNRMLLELLRHLVNV-----CE 512
 Qy 668 HCBE--LLTQALSGRFSNHVHVEYEWRLHDELDDESDDDMDKLPQSENRRREDRPISFYQL 726
 Db 513 HSKENLMT-----SNMGVIFGPTLMRAQEDTVAAMNKKFQ----- 549
 Qy 727 GSNLOQNAVSLAADAANLAKEKORAFWPSILQNETYALLSGSPFPAPQAPSTTSAPP 786
 Db 550 -----NIVVE-----ILIEHFGKIYLG--PPEESA-----APP 575
 Qy 787 LPPRNV---GKVOTASS---ANTLWKTNSVSDGGSRRQ---SSSDP--PAVH 828
 Db 576 VPPPRVTARRHKPITISKRLRLERTVFTYTSLLDSEDEIQHTPNGTNGITTSIEIEPKPPQH 635
 Qy 829 PPLPPLRVSTNPLTPPPPPVAKTPSVMEALSQPSKPAP-----PGISQIRP 876
 Db 636 PKLPIQRSGEDDPGRKSPSRPI-----LDCKLEPCPEVDVGVKLVSLQDGGTKITP 686
 Qy 877 P-----PLPPQPSRLPQ---KKPAPGTDKSTPLTNKGQPRGPVLDLSEALGPIJSNAMVL 929
 Db 687 KATNGPMPSGSGTKTPSFHIKRPA-----RPLAH--HKEGDAD--SFSKVRPGEKPTII 738
 Qy 930 QPPA--PMPKRSQATKLPKRVKALYNCVADNPDELTPS 966
 Db 739 RPPVRPPDPCCRAATPQKPEPKP---DIVAGNAGEITSS 774

RESULT 14

C59431

centaurin, delta 2 [imported] - human

C:Species: Homo sapiens (man)

C:Date: 03-Jun-2002 #sequence revision 03-Jun-2002 #text_change 09-Jul-2004

C:Accession: C59431; D59431

R:Nagase, T.; Ishikawa, K.; Suyama, M.; Kikuno, R.; Miyajima, N.; Tanaka, A.; Kotani, H.

DNA Res. 5, 277-286, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. XI. The completed sequence of the coding sequence of the gene for protein kinase C- δ .
A:Reference number: C59431; MUID:99087487; PMID:9872452
A:Accession: C59431
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1210 <NAG>
A:Cross-references: UNIPROT:Q96L71; GB:NP_056057; PID:g16975484; PIDN:NP_056057.1
R:Miura, K.; Jacques, K.M.; Stauffer, S.; Kubosaki, A.; Zhu, K.; Hirsch, D.S.; Resau, J. Mol. Cell 9, 109-119, 2002
A:Title: ARAP1, A Point of Convergence for Arf and Rho Signaling.
A:Reference number: D59431
A:Accession: D59431
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1210 <NAG>
A:Cross-references: GB:NP_056057; PID:g16975484; PIDN:NP_056057.1

Query Match 5.1%; Score 271; DB 2; Length 1210;
Best Local Similarity 24.1%; Pred. No. 1.4e-05;
Matches 117; Conservative 74; Mismatches 178; Indels 116; Gaps 20;
Qy 211 KKGVDLLQNLKLYFAQCNFFQDGLKAVESLK-----PSTETLSTD-LHTIKQAQDEE 262
Db 107 KRWRVLDTHDYFDNSNDAYSKRISVACISHVAAGDQFVETNNRTTAFRAESDVE 166
Qy 263 RRQLQLRDILKSAQVQKED-SQIROSTAYSLHQPOQNKHEGHTERNGSL-----YKSD 317
Db 167 RKEWQ-----ALQQAWEQARARLSSAYLLGVP-GSEQ--PDRAGSLRFGFNKL 216
Qy 318 GIRKVMQKRCVKVN-----GFLTISHGTANRPPAKLNLITCOVKNPBEKKCPD 367
Db 217 YVAVVGDKVQL-YKNLEBYHLGIGITFIDMSVGNVKE-----VDRRSFD 259
Qy 368 LISHDRTHYQAEDEQEQIWMVSLQNSKEALNNAFNGDDNTGNNIVQELTKELISEV 427
Db 260 LTPYRIFSFSADESEKQEMLEAMQGAIALS-----TSEVARI 301
Qy 428 QRMTGNDVCCGAPDPTWLSNLGILTCIECGIHRELGHVHSPMSQSLTDLVLTSELL 487
Db 302 WAAAPNRFACDCAQPDWASINLCVICKCAGEHGLGAGVKVSLKMDRKWTTL 361
Qy 488 --LAKNIGNAGNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYAR-----K 538
Db 362 IELFLQNGAGNRFAANVPPSEALQPS--SSPSTRCHLEAKYREGKYRYPHLPFGNQ 419
Qy 539 KHADNAKLSICEAVKTRDIF--GLAQYADV-----DLTEKIPLANGHEPDETALH 590
Db 420 BELDKA-----LCAAVTTTDLAETQALLGCGA-GINCFSGDPEAPTPLALAEQAGTLOM 473
Qy 591 LAVRSVDRSTSLHIVDFL-----VQNSGNLDKQTKGST-----ALHY 627
Db 474 EFLNRRTTEVRLDSMKPLEKHYSVLPVTHSHGFLYKTASAGKLQDRRAREEFSRW 533
Qy 628 CCLTD 632
Db 534 CVLGD 538

RESULT 15
F59430
GTPase regulator associated with focal adhesion kinase pp125 [imported] - human
C:Species: Homo sapiens (man)
C:Date: 03-Jun-2002 #sequence revision 03-Jun-2002 #text_change 09-Jul-2004
C:Accession: F59430; G59430; H59430
R:Taylor, J.M.; Macklem, M.M.; Parsons, J.T.
A:Title: Cytoskeletal changes induced by GRAF, the GTPase regulator associated with focal adhesion kinase pp125, are dependent on the tyrosine phosphorylation of GRAF.
A:Reference number: F59430
A:Accession: F59430
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-814 <TAY>

A:Cross-references: UNIPROT:Q9UNAL; GB:NP_055886; PID:g7662208; PIDN:NP_055886.1
R:Borkhardt, A.; Sojensen, S.; Haas, O.A.; Fuchs, U.; Bartelheimer, D.; Loncarevic, I. F., Proc. Natl. Acad. Sci. U.S.A. 97, 9168-9173, 2000
A:Title: The human GRAF gene is fused to MLL in a unique t(5;11)(q31;q23) and both alleles are inactivated in acute myeloid leukemia.
A:Reference number: G59430
A:Accession: G59430
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-814 <BOR>
A:Cross-references: GB:NP_055886; PID:g7662208; PIDN:NP_055886.1
R:Xia, J.H.; Tang, X.X.; Yu, K.P.; Pan, Q.; Dai, H.P.
submitted to GenBank, April 2002
A:Description: Molecular cloning of human oligophrenin-1 like (OPHNIL) gene, complete CTR
A:Reference number: H59430
A:Accession: H59430
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-814 <XTA>
A:Cross-references: GB:NP_055886; PID:g7662208; PIDN:NP_055886.1
Query Match 5.1%; Score 267.5; DB 2; Length 814;
Best Local Similarity 18.2%; Pred. No. 1.2e-05;
Matches 192; Conservative 157; Mismatches 361; Indels 347; Gaps 40;
Qy 34 RNTVAATEALDVRMVLVKMKSVKAINSSGLAHVENEQVQALEKPGNCV-----CR 89
Db 19 RETLKSHEAELDKTKETKELIKDGLSLISALKNLSSAKRFADSLNFKQICGDAETD 78
Qy 90 DDPDLSAFLKFSVFTKELTALFKNLIONMNIISFPDLSLLKGLKLVKGLKPFKDKA 149
Db 79 DEMCIARSLQEFATVLRNLEDERIRMIENASEVLITPLEKPRKEQI-GAAEAKKKYDKE 137
Qy 150 WKDY---ETKITKIEKKEKHAHLGMIRITISGAETAEEMEKERRFPQLQCEYLLKVN 206
Db 138 TEKYCGILEKHLNLSKKCKE-----SQLQADSOVDLVRQ-HFEVSL-EYFVKVQ 186
Qy 207 EIKTKGVLDLQNLKLYFAQCNFFQDGLKAVESLKPSIETLSTDLTHTIKQAQDEERQL 266
Db 187 EVQERKMPFEPVPLLAFLQGLFTFYHGYELAKDFGDKTQITISIQNTRNREFGTRSE 246
Qy 267 IQLRDIKSAQVQKEDSQIROSTAYSLHQPOQNKHEGHTERNGSLYKKSQIRKVMOKR 326
Db 247 ---ESLMKKMKENPLEHKTISPVTMEGYLVQEKRFHFTS-----WKH 287
Qy 327 KCSVNGKPLTIS-----HGTANRPPAKLNLITC-QVKTNPBEKK-CFDLISHDR--TY 375
Db 288 YCTYQDSKQITWVPFDDQKSGKGEDESILKSTREKTDSEIKRFCFDEAVDRPGVI 347
Qy 376 HFQADEQECQIWMVSLQNSKEALNNAFNGDDNTGE---NNIVQELTKELISEVQRMGT 432
Db 348 TMQALSEDRRLWMEAMDG--REPYNKDSQSEGTQAQLDSIGFSIIRKCIHAVETROI 405
Qy 433 NDVCCDCGAPDPTWLSNLGILTCIECGIHRELGHV--YSPMQSLTDLVLTSELLAK 490
Db 406 NE-----QGLYRIVGVNSRVQKLSVLMDPKTASE----- 435
Qy 491 NIGNAGNEIMECCLPADSVK-----PNPGSDMNARKDYITAKYIERRYAR 537
Db 436 -----TETDICAWEIKTITSALKTYRLMPLPGPLMMYQFQSFKAALKENQESR 485
Qy 538 KGHADNAKLSICEAV--KTRDIFGLLQYADVGLTEKIPLANGHEPD-ETALHLAV- 593
Db 486 -----VSEIHSVLRHLPKRNQRLQMLNHLNV-----ANNHQNLMVTANLGV 531
Qy 594 -----RSVDRSTSLHIVDFLVQNSGNLDKQTKGSTALHYCCLTDNAECLKLLRGKASI 647
Db 532 FGPTLLRPEQETVAAMIDKFN-----IVIEILTENHE----- 565
Qy 648 EIANESGTPLDIAKRLKHECEBLLTQALSGRFNSHVHVEYWRLLHEDDESDDMD 707
Db 566 KIFVTVPDPLTNAQ-----LHLSRKSSDSK-- 592
Qy 708 KLPQSENRRDRPIS-FYQLGSLQNSNAVSLARDAANLAKEKQAFMPISILQNETYGA 766

```
Db 593 --PPS---CSRPLTLFHTVQSTEQ-----EQRNSIINSLE----- 625
Qy 767 LSGSPPPAQAAPSTTSAPPLPPRNVGVKQTASSANTLWKTNVSVDGGSRQ--RSSSDP 824
Db 626 -----SVSSNP-----NSILNSSSSSLQPNMNSSD- 649
Qy 825 PAVHPPLPPLRVSTNPLTPTPPPPVAKTPSVMEALSQPSKPAPPGISQIRPPP----- 878
Db 650 ----PDLAVVRPTRPNSLPPNPSPTSPLSPS-WPMFSAPSSPMPSTSSSDSSPVRSVAG 704
Qy 879 -----LPPQPPSRLPQKKP--APGTDKSTPLTNKGQP 908
Db 705 FWFSSVAAVVLSARSSLHAVFSLVNFVCPHNLHLLFDRPEEAVHEDSSTPF----- 758
Qy 909 RGPVDLSATEALGPLSNAMVLQPPAPMPRKSQATKPKRVKALYNCVADNPDELTFSEG 968
Db 759 -----RKAKALYACKAEHDSLSFTAG 780
Qy 969 DVI-IVDGEEDQEWIGHIDGDPGRKGAPPVSFVHFI 1004
Db 781 TVFDNVHPSPQEPGWLEGTLN---GKTGLIPENYVEFL 814
```

Search completed: August 4, 2005, 08:27:21
Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:07:43 ; Search time 111 Seconds
(without alignments)
4641.006 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHEDYKA.....DGDPRKAGPVPVHFIAID 1006

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5253	99.6	1006	1	DDF2_HUMAN
2	4727	89.6	970	2	Q6A074
3	4215.5	79.9	1029	2	Q6NRK0
4	3877.5	73.5	809	2	Q66JN2
5	3867	73.3	784	1	DDF2_MOUSE
6	3013.5	57.1	1129	1	DDF1_BOVIN
7	3000.5	56.9	1147	1	DDF1_MOUSE
8	2465.5	46.8	956	1	DDF1_HUMAN
9	2071.5	39.3	903	2	Q8TDY4
10	2067.5	39.2	903	2	Q6P9F4
11	1499.5	28.4	1015	2	Q7QJY3
12	1380	26.2	935	2	Q6AWJ6
13	1075.5	20.4	208	2	Q8N282
14	878	16.7	307	2	Q6DEH1
15	844	16.0	762	2	Q9V367
16	655.5	12.4	439	2	Q8BM35
17	631	12.0	424	2	Q86UY1
18	587.5	11.1	807	2	Q6ZKX5
19	583	11.1	307	2	Q9NXH7
20	582	11.0	407	2	Q9NKK2
21	575.5	10.9	778	1	Q6IVG4
22	573.5	10.9	778	1	CEB2_HUMAN
23	543	10.3	858	2	Q86UT3
24	533	10.1	759	1	CEB5_HUMAN
25	523	9.9	833	2	Q6NXL5
26	515	9.8	740	2	Q8K2H4
27	509.5	9.7	775	2	Q8SMX5
28	509	9.7	740	1	CEB1_HUMAN
29	504.5	9.6	843	2	Q86JDI
30	490.5	9.3	768	2	Q9FIT8
31	487.5	9.2	826	2	Q9XXH8

32	487.5	9.2	826	2	Q9XZQ1
33	485.5	9.2	776	2	Q9C6C3
34	481	9.1	760	2	Q6H462
35	479.5	9.1	742	2	Q9XXH9
36	475	9.0	742	2	Q9XZQ2
37	466.5	8.8	519	2	Q6PA21
38	466	8.8	759	2	Q9LYU6
39	455.5	8.6	799	2	Q7QFM2
40	449.5	8.5	552	2	Q8CI47
41	447	8.5	202	2	Q6PJ46
42	447	8.5	753	2	Q6ZD39
43	445	8.4	558	2	Q9SR80
44	441	8.4	487	2	Q6GN75
45	439	8.3	828	2	Q9VCQ6

ALIGNMENTS

RESULT 1

ID	DDF2_HUMAN	STANDARD;	PRT;	1006 AA.
AC	O43150;			
DT	29-MAR-2004 (Rel. 43, Created)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Development and differentiation-enhancing factor 2 (Pylk C-terminus associated protein) (PAP) (Paxillin-associated protein with ARFGAP activity 3) (PAG3).			
GN	Name=DDEF2; Synonym=KIAA0400;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
EN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Brain;			
FX	MEDLINE=98116655; PubMed=9455477;			
RA	Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. VIII. 78 new cDNA clones from brain which code for large proteins in vitro."			
RT	DNA Res. 4:307-313(1997).			
RL	[2]			
RN	SEQUENCE FROM N.A. (ISOFORM 2).			
RP	TISSUE=Placenta;			
RC	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
FX	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Villalobos D., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Richards S.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Whiting J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A., Blakeley R.C., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).			
EN	[3]			
RP	FUNCTION, PHOSPHORYLATION, INTERACTION WITH ARF1; ARF5; PTK2B			
RP	AND SRC, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE SPLICING.			
RP	PubMed=10022920;			

RA Andreev J., Simon J.-P., Sabatini D.D., Kam J., Plowman G.,
 RA Randazzo P.A., Schllessinger J.,
 RT "Identification of a new Fyk2 target protein with Arf-GAP activity.",
 RL Mol. Cell. Biol. 19:2338-2350(1999).
 [4]
 RN FUNCTION, MUTAGENESIS OF CYS-436, SUBCELLULAR LOCATION, AND
 RP INTERACTION WITH PKN.
 RX PubMed=1074932;
 RA Kondo A., Hashimoto S., Yano H., Nagayama K., Mazaki Y., Sabe H.;
 RT "A new paxillin-binding protein, PAG3/Papalpa/KIAA0400, bearing an
 RT ADP-ribosylation factor GTPase-activating protein activity, is
 RT involved in paxillin recruitment to focal adhesions and cell
 RT migration.";
 RL Mol. Biol. Cell 11:1315-1327(2000).
 [5]
 RN FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH ARF6 AND ACTIN
 RP FILAMENTS.
 RX PubMed=11304556;
 RA Uchida H., Kondo A., Yoshimura Y., Mazaki Y., Sabe H.;
 RT "PAG3/Papalpa/KIAA0400, a GTPase-activating protein for ADP-
 RT ribosylation factor (ARF), regulates ARF6 in Fcgamma receptor-mediated
 RT phagocytosis of macrophages.";
 RL J. Exp. Med. 193:955-966(2001).
 CC -|- FUNCTION: Activates the small GTPases ARF1, ARF5 and ARF6.
 CC Regulates the formation of post-Golgi vesicles and modulates
 CC constitutive secretion. Modulates phagocytosis mediated by Fc
 CC gamma receptor and ARF6. Modulates PAXN recruitment to focal
 CC contacts and cell migration.
 CC -|- SUBUNIT: Binds PKN, ARF1, ARF5, ARF6, PTK2B and SRC.
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic and periphera membrane protein
 CC associated with the plasma membrane and with Golgi stacks.
 CC Colocalizes with F-actin and ARF6 in phagocytic cups.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=PAPAlpha;
 CC IsoId=043150-1; Sequence=Displayed;
 CC Name=2; Synonyms=PAPbeta;
 CC IsoId=043150-2; Sequence=VSP_009722;
 CC -|- TISSUE SPECIFICITY: Detected in heart, brain, placenta, kidney,
 CC monocytes and pancreas.
 CC -|- INDUCTION: Up-regulated during monocyte maturation.
 CC -|- DOMAIN: The conserved Arg-464 in the Arf-GAP domain probably
 CC becomes part of the active site of bound small GTPases and is
 CC necessary for GTP hydrolysis.
 CC -|- PM: Phosphorylated on tyrosine residues by SRC and PTK2B.
 CC -|- SIMILARITY: Contains 2 ANK repeats.
 CC -|- SIMILARITY: Contains 1 Arf-GAP domain.
 CC -|- SIMILARITY: Contains 1 PH domain.
 CC -|- SIMILARITY: Contains 1 SH3 domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AB007860; BAA23696.2; ALT_INIT.
 DR EMBL; BC063308; AAH63308.1; -.
 DR HSSP; O60631; 1GBQ.
 DR Genew; HGNC:2721; DDEF2.
 DR MIM; 603817; -.
 DR GO; GO:0008047; F:enzyme activator activity; TAS.
 DR ProDom; P000066; SH3; 1.
 DR PROSITE; PS00297; ANK REP REGION; 1.
 DR PROSITE; PS00088; ANK REPEAT; 1.
 DR PROSITE; PS00115; ARFGAP; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW Alternative splicing; ANK repeat; Coiled coil; Golgi stack;
 KW GTPase activation; Metal-binding; Phosphorylation; Repeat; SH3 domain;
 Zinc.

FT	DOMAIN	305	397	PH.
FT	DOMAIN	421	543	Arf-GAP.
FT	REPEAT	584	616	ANK 1.
FT	REPEAT	620	649	ANK 2.
FT	DOMAIN	944	1006	SH3.
FT	DOMAIN	256	283	Coiled coil (Potential).
FT	DOMAIN	729	752	Coiled coil (Potential).
FT	DOMAIN	771	936	Pro-rich.
FT	VARSPLIC	795	840	VQTASSANTLWKINSVSDGSRQRSSSDPPAVHPPLPPLR
FT				VTSTN -> D (in isoform 2).
FT				/FTId=VSP_009722.
FT	MUTAGEN	436		C->A: Loss of Arf-GAP activity.
FT	CONFLICT	86		C -> R (in Ref. 2).
FT	CONFLICT	748		E -> D (in Ref. 2).
SQ	SEQUENCE	1006 AA;	111650 MW;	6A213517DCD99E1B CRC64;
	Query Match	99.6%;	Score 5253;	DB 1; Length 1006;
	Best Local Similarity	99.7%;	Pred. No. 1.2e-213;	
	Matches 1003;	Conservative 0;	Mismatches 3;	Indels 0; Gaps 0;
Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCHNTVAATIEEALDVRMVLVYKMKSKVKA	60	
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCHNTVAATIEEALDVRMVLVYKMKSKVKA	60	
Qy	61	INSSGLAHVNEEOYTQALEKFGGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLQNMN	120	
Db	61	INSSGLAHVNEEOYTQALEKFGGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLQNMN	120	
Qy	121	NIISFPDLSLLKGLKGVKGDLLKPKFPAKWDYETKITKEKEKEHAKLHGMIRTEISG	180	
Db	121	NIISFPDLSLLKGLKGVKGDLLKPKFPAKWDYETKITKEKEKEHAKLHGMIRTEISG	180	
Qy	181	AETAEEMEKERRFPQLOMCEYLLKWEIKIKKGVDDLLQNLKIKYHQAQCNFQDGLKAVES	240	
Db	181	AETAEEMEKERRFPQLOMCEYLLKWEIKIKKGVDDLLQNLKIKYHQAQCNFQDGLKAVES	240	
Qy	241	LKPSIETLSTDLHTIKQAQDEERQLIQRLDKLSALQVEKESQIRQSTAYSLHPOQG	300	
Db	241	LKPSIETLSTDLHTIKQAQDEERQLIQRLDKLSALQVEKESQIRQSTAYSLHPOQG	300	
Qy	301	NKEHGTERNGSLYKKSQDGIIRKVMQKRCVKNQGLTISHGTANRPPAKLNLTCQVKTNP	360	
Db	301	NKEHGTERNGSLYKKSQDGIIRKVMQKRCVKNQGLTISHGTANRPPAKLNLTCQVKTNP	360	
Qy	361	BEKKCFDLISHDRTYHFQAEDEBQCIQWMSVLQNSKEEALNNAFKGDDNTGNNIVQELT	420	
Db	361	BEKKCFDLISHDRTYHFQAEDEBQCIQWMSVLQNSKEEALNNAFKGDDNTGNNIVQELT	420	
Qy	421	KEIISEVQRMVGTNDVCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMSQSLTLDV	480	
Db	421	KEIISEVQRMVGTNDVCCGAPDPTWLSTNLGILTCIECSGIHRELGVHYSRMSQSLTLDV	480	
Qy	481	LGTSSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNAKDYITAKYIERRYARKKH	540	
Db	481	LGTSSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNAKDYITAKYIERRYARKKH	540	
Qy	541	ADNAAKLHSLCEAVKTRDIFGLLOAYADGVDLTKIPLANGHEPDETHALHVAHSVDRTS	600	
Db	541	ADNAAKLHSLCEAVKTRDIFGLLOAYADGVDLTKIPLANGHEPDETHALHVAHSVDRTS	600	
Qy	601	LHVIDELVQNSGNDKQTKGKSTALHYCCLTDNAECLLRLGKASIEANESGETPLDI	660	
Db	601	LHVIDELVQNSGNDKQTKGKSTALHYCCLTDNAECLLRLGKASIEANESGETPLDI	660	
Qy	661	AKRLKHEHCEBELLTQALSGRFSNHSVHYEYEWRLHEDLDESDDMDKELQSPENRRDRP	720	
Db	661	AKRLKHEHCEBELLTQALSGRFSNHSVHYEYEWRLHEDLDESDDMDKELQSPENRRDRP	720	
Qy	721	ISFYQLGNSQLQSNVSLARDAAANLAKEKQAFMPSILQNETYCALLISGSPPPQAPAPS	780	
Db	721	ISFYQLGNSQLQSNVSLARDAAANLAKEKQAFMPSILQNETYCALLISGSPPPQAPAPS	780	
Qy	781	TTSAPPLPRNVGVQVQTASSANTLWKNTNSVSDGSRQRSSSDPPAVHPPLPPLRVVTSTN	840	

```
Db 781 TTSAPPLPRNVKQVQTASSANTLWKTNSVSDGSGRQSSDDPPAVHPPLPPLRVSTN 840
Qy 841 PLTPPPPPVAKTPSVMEALSPKAPPAGISQIRPPPLPPOPSPRLPQKPPAGCTDKST 900
Db 841 PLTPPPPPVAKTPSVMEALSPKAPPAGISQIRPPPLPPOPSPRLPQKPPAGCTDKST 900
Qy 901 PLTNKGQPRGVDLSATBALGSLNAMVLQPPAPMPKRSQATKLPKRVKALYNCAVADNP 960
Db 901 PLTNKGQPRGVDLSATBALGSLNAMVLQPPAPMPKRSQATKLPKRVKALYNCAVADNP 960
Qy 961 DELTSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 1006
Db 961 DELTSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 1006

RESULT 2
Q6A074 PRELIMINARY; PRT; 970 AA.
ID Q6A074
AC Q6A074;
DT 25-OCT-2004 (TREMELrel. 28, Created)
DT 25-OCT-2004 (TREMELrel. 28, Last sequence update)
DE MKTAA0400 protein (fragment).
GN Name=mkTAA0400;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Saga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-Fractionated Libraries.";
RL DNA Res. 11:205-218(2004).
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC EMBL; AK172944; BAD3222.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR01036; PH_related.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3_2.
DR Pfam; PF00023; Ank; 3.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3_1; 1.
DR Pfam; PF07653; SH3_2; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00405; REVINTACTING.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; SH3 domain.
FT NON_TER 1
SQ SEQUENCE 970 AA; 107768 MW; 11BAAL16234DFECL CRC64;
```

Query Match 89.6%; Score 4727; DB 2; Length 970;

RESULT 3

```
Best Local Similarity 90.0%; Pred. No. 1.7e-191;
Matches 908; Conservative 18; Mismatches 29; Indels 54; Gaps 3;
Qy 1 MPDOIISSEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVMVLYKMKKS VKA 60
Db 13 MPDOIISSEFVAETHEDYKAPTASFTTRTAQCRNTVAABEALDVMVLYKMKKS VKA 72
Qy 61 INSSGLAHVENEQYQTALEKFGGNCVCRDPDGLGSAPFKFSVFTKELTALFKNLIONNN 120
Db 73 INISGLAHVENEQYQTALEKFGGNCVCRDPDGLGSAPFKFSVFTKELTALFKNLIONNN 132
Qy 121 NIISFPLDSLKGLDKGVKGLKKPDKAWKDYETKTKIEKKEKHAHLGHMIRTEISG 180
Db 133 NIISFPLDSLKGLDKGVKGLKKPDKAWKDYETKTKIEKKEKHAHLGHMIRTEISG 192
Qy 181 AEIAEEMEKERRFFOLOMCEYLLKNEIKIKKGVLDLQNLIKYFHAQCNFFODGLKAVES 240
Db 193 AEIAEEMEKERRFFOLOMCEYLLKNEIKIKKGVLDLQNLIKYFHAQCNFFODGLKAVES 252
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEQKE---DSQIROSTAYS LHQ 297
Db 253 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEQKESRRDSQIROSTAYS LHQ 312
Qy 298 PQGNKEHGTENGSLYKKSQGIKRWQKXKCSVKNGLFTISHGTANRRPPAKNLILTCQVK 357
Db 313 PQGNKEHGTENGSLYKKSQGIKRWQKXKCSVKNGLFTISHGTANRRPPAKNLILTCQVK 372
Qy 358 TNPEKKCFDLISHDRTYHFOAEDEQECQIWNMSVLONSKEALNNAFKGDDNTGNNIVQ 417
Db 373 TNPEKKCFDLISHDRTYHFOAEDEQECQIWNMSVLONSKEALNNAFKGDDNTGNNIVQ 432
Qy 418 ELTKIEIISVQRMGTGNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGHVHYSMQSLT 477
Db 433 ELTKIEIISVQRMGTGNDVCCDGPDPPTWLTSTNLGILTCIECSGIHRELGHVHYSMQSLT 492
Qy 478 LDVLGTSELLAKNIGNAGFNEIMECCLPDSDVKNPNSGDMNARKDVIYAKIYERRYAR 537
Db 493 LDVLGTSELLAKNIGNAGFNEIMECCLPDSDVKNPNSGDMNARKDVIYAKIYERRYAR 552
Qy 538 KKHADNAAKHSLCEAVKTRDIFGLQAYAGVDLTETKIPLANGHEPDETALHLAVRSD 597
Db 553 KKHADTAAKHSLCEAVKTRDIFGLQAYAGVDLTETKIPLANGHEPDETALHLAVRSD 612
Qy 598 RTSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETP 657
Db 613 RTSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETP 672
Qy 658 LDIAKRLKHECEBELLTQALSGRFSNVHVEYEWRLHEDLDESDDDDDEKLOPSNRRE 717
Db 673 LDIAKRLKHECEBELLTQALSGRFSNVHVEYEWRLHEDLDESDDDDDEKLOPSNRRE 732
Qy 718 DRPISFYQLGSLQNSNAVSLARDAANLAKESQAFMPSILQNETYGALLSGSPPPAOPA 777
Db 733 DRPVSFYQLGSSQFOSNAVSLARDTANUTKQKQKGFSGSILQNETYGALLSGSPSSQSI 792
Qy 778 APSTTSAPPLPRNVKQVQTASSANTLWKTNSVSDGSGRQSSDDPPAVHPPLPPLRV 837
Db 793 PPSITTSAPPLPRNVGK----- 809
Qy 838 STNPLTPTPPPPVAKTPSVMEALSPKAPPAGISQIRPPPLPPOPSPRLPQKPPAGTGD 897
Db 810 --DPLTTTPPPVAKTSGTLEAMNPQSKSQSGTSQSKPPPLPPOPSPRLPQKPPAGTGD 867
Qy 898 KSTPLTNKGQPRGVDLSATBALGSLNAMVLQPPAPMPKRSQATKLPKRVKALYNCAV 957
Db 868 KPTPLTNKGQPRGP-----EASGPLSNAMALQPPAPMPKRSQATKLPKRVKALYNCAV 921
Qy 958 DNPDELTFSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 1006
Db 922 DNPDELTFSEGDVLIIDGEEDEQEWIGHIDGDPGRKGAPFVSFVHFAD 970
```


Q6NRK0
 ID Q6NRK0 PRELIMINARY; PRT; 1029 AA.
 AC Q6NRK0;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE MGC83760 protein.
 GN Name=MGC83760;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Oocytes;
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny D.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Fahey J.J., Helton E., Kettner M.M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Oocytes;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Oocytes;
 RA Klein S., Strausberg R.;
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Contains 1 PH domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 DR EMBL; BC070750; AAH70750.1; -.
 DR HSSP; P19878; 1KAU.
 DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
 DR InterPro; IPRO02110; ANK.
 DR InterPro; IPRO01164; ArfGAP.
 DR InterPro; IPRO01849; PH.
 DR InterPro; IPRO11036; PH related.
 DR InterPro; IPRO01452; SH3.
 DR Pfam; PF00023; Ank; 2.
 DR Pfam; PF01412; ArfGAP; 1.
 DR Pfam; PF00169; PH; 1.
 DR PRINTS; PR01415; ANKYRIN.
 DR PRINTS; PR00405; REVINTRACTING.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00248; ANK; 3.
 DR SMART; SM00105; ArfGAP; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REGION; 1.

DR PROSITE; PS50115; ARFGAP; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW ANK repeat; SH3 domain.
 SQ SEQUENCE 1029 AA; 114129 MW; BCSFCS7D0044A8C5 CRC64;
 Query Match
 Best Local Similarity 79.9%; Score 4215.5; DB 2; Length 1029;
 Matches 806; Conservative 85; Mismatches 113; Indels 27; Gaps 7;
 QY 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
 DB 1 MPDLISVSEFVCTNEDYKSPASNTNRLSQCRNTVAIDAELEVDHTILNKKMKSVKA 60
 QY 61 INSSGLAHVENEEOYTOALEKFGNGCYCRDDPDGLGSAPLKESVFTKELTALFKNLIQNMN 120
 DB 61 IHTSGLTHVENEEOYTOALEKFGNGCYCRDDPDGLGTAFLEKFSVFTKELTALFKNLVQNMN 120
 QY 121 NIIFPPLDSLKGLKGVKGLKKPPDKAWDKYETKTKIEKEKEHAKHGMIRTEISG 180
 DB 121 NIITFPDLSLLKGLKGVKGLKKPPDKAWDKYETKTKIEKEKEHAKHGMIRTEISG 180
 QY 181 AEIAEEMEKERRFPQLOMCEYLLKKNBIKIKGGVDLLQNLIKYFHAQCNPFDGLKAVES 240
 DB 181 AEIAEEMEKERRFPQLOMCEYLLKKNBIKIKGGVDLLQNLIKYFHAQCNPFDGLKAVET 240
 QY 241 LKPSIETLSTDLHTIKQADEERQLIQLRDILKSALQVEQKESQIRQSTAYSILHQPOG 300
 DB 241 LKPSIEKLSTDLHTIKQVDEERKQLSRLDKAALQVEQKESQIRQSTAYSILHQPOG 300
 QY 301 NKEHGTRNGSLYKKSIGIRKWKVKRCSVQNGFLTTISHTGANPPAKLNLLTTCQVKTNP 360
 DB 301 NKEHGTRDGYLYKKSIGIRKWKVKRCSVQNGFLTTISHTGANPPAKLNLLTTCQVKTNP 360
 QY 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSVLQNSKEEALNAPFKGDDNTGNNIVBELT 420
 DB 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSVLQNSKEEALNAPFKGDDNTGNNIVBELT 420
 QY 421 KEIISVQRMVGNVDCDCGAPDPTWLTNLTGILTCIECSGIHRELGVHYHSPMSLTLDV 480
 DB 421 KEIISDIQRVPGNVDCDCGASDPTWLTNLTGILTCIECSGIHRELGVHYHSPMSLTLDV 480
 QY 481 LGTSELLAKNIGNAGFNEIMECCLPADSDSKPTPCSDMNAKDYITAKYIERYARKKH 540
 DB 481 LGTSELLAKNIGNSGFNEVMEACLPAEDSLKPTPCSDMNAKDYITAKYIERYARKKY 540
 QY 541 ADNAAKLHSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVSRVDRTS 600
 DB 541 HDNAKSLHGLCDAVKSRDIPALVHLHAEGVDLTETAIPLANGHEQDETALHLAVSRVDRNS 600
 QY 601 LHIYDFLVQNSGNDLKQTKGSTALHVCCLTDNAECLKLLRGKASTEIANESGETPLDI 660
 DB 601 LHIYDFLVQNSGNDLKQTKGSTALHVCCLTDNAECLKLLRGKASTDIANDAGETPLDI 660
 QY 661 AKRLKHEHEBELLTQALSGRENSHVHYEYEWRLHEDLDESDDDMDKEL-OPSENRREDR 719
 DB 661 AKRLKHAQCEDLIIQAMNGRFNSHVHYEYEWRLHVDLDESDDDVDEKQKQSPIRREER 720
 QY 720 PISFYQLSGNQLQSNVSLARDAANLAKERAKRAMPISQILQNETYCALLSGSPPPAQAAP 779
 DB 721 PVSFYQLGNGQPSVACLAREAAIAKDKPRNFPLNIIQNETYGAIVGNSNSPGNAAGP 780
 QY 780 STTSAPLPPRNVGK-----QTASSANTLWKTNVSVDGSGRQRSSDPPAVHPP 830
 DB 781 ATTNAPLPPRNNTIKVHSTIFGSSIHPTPSNTPAWKASSISAESGRQRSSDPPSIQPP 840
 QY 831 LPPLRVSTNELTPTPPPVAKTSPVMEALSQSPKAPP---GISQRPPLPPPOPPSRL 887
 DB 841 VPPLRVST-ATNPVPPPPVTKTASVMEALSQHSKQSPPPVPLNKPVPVPPVQPANRP 899
 QY 888 PQKPAFQTDKSTPLTNKGQPRGPVDLSATEAL-----GPILSNAMV-LQPPAPM 935
 DB 900 SHKRPAPGIDKPA-ATTGKPPPALVTASLPAADFPSETTSSQPGPKSLTPSPAPM 958

Db 601 SSQIPSTTSAPPLPRNVGVK----- 623

Qy 833 PLRVSTNPLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPPQPPSRLPQKPP 892

Db 624 -----DPLTTTPPPVAKTSGTLEAMNQPSKSSQPGTSQSKPPPLPPQPPSRLPQKPP 676

Qy 893 APGTDKSTPLTNKQPGPRVDLSATEALGPLSNAMVLOPPAPMPKRSQATKLPKRVKAL 952

Db 677 ASGADKPTPLTNKQPGPRP-----EASGPLSNAMALOPPAPMPKRSQATKLPKRVKAL 730

Qy 953 YNCVADNDELTFSEGVDVIIVDGEDQEWLGHIDGDPGRGAPVPSFVHFIAID 1006

Db 731 YNCVADNDELTFSEGVDVIIVDGEDQEWLGHIDGEPGRGAPVPSFVHFIAID 784

RESULT 6

DDFL_BOVIN

ID_DDF1_BOVIN STANDARD; PRT; 1129 AA.

AC O97902;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent Arf1 GTPase-activating protein (PIP2-dependent Arf1 GAP) (ADP-ribosylation factor-directed GTPase-activating protein 1) (Arf GTPase-activating protein 1) (Development and differentiation-enhancing factor 1) (DEF-1).

DE 1) (Differentiation-enhancing factor 1) (DEF-1).

GN Name=DEF1; Synonyms=ASAP1;

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

OC NCBI_TaxID=9913;

OX [1]

RN SEQUENCE FROM N.A., CHARACTERIZATION, AND SEQUENCE OF 80-101; 303-324; 550-558; 804-826; 839-854 AND 1039-1049.

RP TISSUE=Brain;

RX MEDLINE=99147067; PubMed=10022919;

RA King F.J., Hu E., Harris D.F., Sarraf P., Spiegelman B.M., Roberts T.M.;

RA "DEF-1, a novel src SH3 binding protein that promotes adipogenesis in fibroblastic cell lines.";

RT Mol. Cell. Biol. 19:2330-2337(1999).

RL [2]

RN PARTIAL SEQUENCE FROM N.A., AND SEQUENCE OF 61-65; 130-133; 424-427 AND 502-525.

RP TISSUE=Brain;

RC MEDLINE=99038209; PubMed=9819391;

RX Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A., Randazzo P.A.;

RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that associates with and is phosphorylated by Src.";

RL Mol. Cell. Biol. 18:7038-7051(1998)

CC -!- FUNCTION: Possesses phosphatidylinositol 4,5-bisphosphate-dependent GTPase-activating protein activity for Arf1 (ADP ribosylation factor 1) and Arf5 and a lesser activity towards Arf6. May coordinate membrane trafficking with cell growth or actin cytoskeleton remodeling by binding to both SRC and PIP2 (By similarity). May function as a signal transduction protein involved in the differentiation of fibroblasts into adipocytes and possibly other cell types.

CC -!- ENZYME REGULATION: Activity stimulated by phosphatidylinositol 4,5-bisphosphate (PIP2) (By similarity).

CC -!- SUBUNIT: Homodimer. Interacts with SRC and CRK (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-associated; partially (By similarity).

CC -!- TISSUE SPECIFICITY: Ubiquitous.

CC -!- DOMAIN: The PH domain most probably contributes to the phosphoinositide-dependent regulation of ADP ribosylation factors (By similarity).

CC -!- PTM: Phosphorylated on tyrosine residues by SRC (By similarity).

CC -!- SIMILARITY: Contains 2 ANK repeats.

CC -!- SIMILARITY: Contains 1 Arf-GAP domain.

CC -!- SIMILARITY: Contains 1 C4-type zinc finger.

CC -!- SIMILARITY: Contains 1 PH domain.

CC -!- SIMILARITY: Contains 1 SH3 domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch)

CC -----

CC EMBL; AF112886; AAD19965.1; -.

DR HSSP; O89100; 1OEB.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR001164; hRIP like.

DR InterPro; IPR000108; Neu_Cyt_fact_2.

DR InterPro; IPR001849; PH_Cyt_2.

DR InterPro; IPR001452; PH_related.

DR Pfam; PF00023; Ank; 2.

DR Pfam; PF00169; PH; 1.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00499; P67PHOX.

DR PRINTS; PR00405; REVINTRACTNG.

DR PRODOM; PD000066; SH3; 1.

DR SMART; SM00248; ANK; 2.

DR SMART; SM00105; ArfGAP; 1.

DR SMART; SM00233; PH; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS00297; ANK REP REGION; 1.

DR PROSITE; PS00088; ANK REPEAT; 2.

DR PROSITE; PS00115; ArfGAP; 1.

DR PROSITE; PS00003; PH DOMAIN; 1.

DR PROSITE; PS00002; SH3; 1.

KW ANK repeat; Direct protein sequencing; GTPase activation; Membrane;

KW Repeat; SH3 domain; Zinc-finger.

FT DOMAIN 327 419 PH.

FT DOMAIN 442 565 Arf-GAP.

FT ZN FING 457 480 C4-type.

FT REPEAT 603 635 ANK 1.

FT REPEAT 639 668 ANK 2.

FT DOMAIN 786 1059 Pro-rich.

FT DOMAIN 1067 1129 SH3.

FT CONFLICT 65 65 D -> T (in Ref. 2; AA sequence).

SQ SEQUENCE 1129 AA; 125381 MW; C1576C2EAC0ACD8 CRC64;

Query Match 57.1%; Score 3013.5; DB 1; Length 1129;

Best Local Similarity 55.3%; Pred. No. 3.9e-119;

Matches 627; Conservative 129; Mismatches 225; Indels 153; Gaps 23;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCNTVAIAIEALDVRMLVKKKSVKA 60

Db 21 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCNTVAIAIEALDVRMLVKKKSVKA 80

Qy 61 INSSGLAHVNEEOYTQALEKFGGNCVCRDDPDGLGSAPLKFPSVFTKELTALPKNLQNMN 120

Db 81 IYNSGDHVQNEENYAQVLDFGNSFLSRDNPDLGTAFVKESTLTKESTLKLKLLQGLS 140

Qy 121 NIISFPLDLSLLKGLKGVKDLKKPDKAMKDYETKITIKBEKKEHAKLHGMIRTEISG 180

Db 141 HNVFTLDSLILKGLKGVKDLKKPDKAMKDYETKITIKBEKKEHAKLHGMIRTEITG 200

Qy 181 AEIAEEMEKERRPQLOWCEYLLKKNVETIKKGVVDLLQNLIKYHAQCNFPQDGLKAVES 240

Db 201 AEIAEEMEKERRPQLOWCEYLLKKNVETIKKGVVDLLQNLIKYHAQCNFPQDGLKAVES 260

Qy 241 LKPSIETLSTDLHTIKQAQDEERQLQLRDLKLSALQVEQKE--DSQIRQSTAYSLSHQ 297

Db 261 LKQVIEKLADLYNIKQTQDEERQLQLRDLKLSALQVEQKE--DSQIRQSTAYSLSHQ 319

Qy 298 PQGNKEHGTERRNGSLYKKSDGIRKVMQKRCVKNVGLTISHGTANRPPAKNLNLTQGVK 357

CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Name=1; Synonyms=SHAG1a, ASAP1a;
 CC IsoId=Q9QWY8-1; Sequence=Displayed;
 CC Name=2; Synonyms=SHAG1b, ASAP1b;
 CC IsoId=Q9QWY8-2; Sequence=VSP_008366;
 CC Name=3;
 CC IsoId=Q9QWY8-3; Sequence=VSP_008366;
 CC Name=4;
 CC IsoId=Q9QWY8-4; Sequence=VSP_008367;
 CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined but a most
 CC abundant expression was found in the testis, brain, lung and
 CC spleen. A heightened expression was seen in the adipose tissue
 CC from obese (ob) and diabetic (db) animals.
 CC -!- DOMAIN: The PH domain most probably contributes to the
 CC phosphoinositide-dependent regulation of ADP ribosylation factors.
 CC -!- PTM: Phosphorylated on tyrosine residues by SRC.
 CC -!- SIMILARITY: Contains 2 ANK repeats.
 CC -!- SIMILARITY: Contains 1 Arf-GAP domain.
 CC -!- SIMILARITY: Contains 1 C4-type zinc finger.
 CC -!- SIMILARITY: Contains 1 PH domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF075461; AAC98349.1; -;
 CC EMBL; AF075462; AAC98350.1; -;
 CC EMBL; AK122477; BAC65759.1; -;
 CC EMBL; BC002201; AAH02201.1; ALT_INIT.
 CC EMBL; BC048818; AAH48818.1; ALT_INIT.
 CC EMBL; U92478; AAB82338.1; -;
 CC PIR; T42627; T42627.
 CC HSSP; P15891; IJ08.
 CC MGD; MGI:1342335; DDef1.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR001164; hRIP_like.
 CC InterPro; IPR000108; Neu_cyt_fact_2.
 CC InterPro; IPR001849; PH.
 CC InterPro; IPR011036; PH related.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00023; Ank; 3.
 CC Pfam; PF01412; ArfGap; 1.
 CC Pfam; PF00169; PH; 1.
 CC Pfam; PF00018; SH3; 1.
 CC PRINTS; PR00499; P67PHOX.
 CC PRINTS; PR00405; REVINTRACTING.
 CC ProDom; PD000066; SH3; 1.
 CC SMART; SM00248; ANK; 2.
 CC SMART; SM00105; ArfGap; 1.
 CC SMART; SM00233; PH; 1.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS0297; ANK_REPEAT; 1.
 CC PROSITE; PS50088; ANK_REPEAT; 2.
 CC PROSITE; PS50115; ARFGAP; 1.
 CC PROSITE; PS50003; PH_DOMAIN; 1.
 CC PROSITE; PS50002; SH3; 1.
 CC Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;
 CC SH3 domain; zinc-finger.
 CC DOMAIN 339 431 PH.
 CC DOMAIN 454 577 Arf-GAP.
 CC ZN_FING 469 492 C4-type.
 CC REPEAT 615 647 ANK 1.
 CC REPEAT 651 680 ANK 2.
 CC DOMAIN 798 1011 Pro-rich.
 CC DOMAIN 1085 1147 SH3.
 CC VARSPLIC 304 315 Missing (in isoform 3).
 CC /FTid=VSP_008366.

FT	VARSPLIC	304	318	Missing (in isoform 4). /FTid=VSP_008367.
FT	VARSPLIC	816	872	Missing (in isoform 2). /FTid=VSP_008368.
FT	MUTAGEN	811	811	R->A: Significant reduction in binding to SRC and CRK and loss of phosphorylation. Loss of binding and phosphorylation; when associated with A-910 and A-913.
FT	MUTAGEN	910	910	P->A: Significant reduction in binding to SRC and CRK and decrease in phosphorylation; when associated with A- 913. Loss of binding and phosphorylation; when associated with A-811 and A-913.
FT	MUTAGEN	913	913	P->A: Significant reduction in binding to SRC and CRK and decrease in phosphorylation; when associated with A- 910. Loss of binding and phosphorylation; when associated with A-811 and A-910.
FT	CONFLICT	654	654	T -> S (in Ref. 4).
FT	CONFLICT	879	879	S -> L (in Ref. 3 and 4).
FT	CONFLICT	1051	1051	R -> I (in Ref. 4).
SQ	SEQUENCE	1147	127395	AA; 1A08321C491B4609 CRC64;

Query Match 56.9%; Score 3000.5; DB 1; Length 1147;
 Best Local Similarity 53.9%; Pred. No. 1.4e-118;
 Matches 621; Conservative 128; Mismatches 232; Indels 171; Gaps 19;

Qy	1	MPDQISVSEFAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLVYMKKSVKA	60
Db	21	MPDQISVSEFAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLVYMKKSVKA	80
Qy	61	INSSGLARHVEEYOYTAQLEKFGGNCVCRDDPDGLGSAPLKFSVFTKELTALFKNLQNN	120
Db	81	LYNSGDHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFVKEFTLTKELSTLKNLQGLS	140
Qy	121	NIISFPLDLLKGLDKGVKGLKTPDKAWKDYETKTKISEKEKEHAKLHGMIRTEISG	180
Db	141	HNVIPTLDSLKGLDKGVKGLKTPDKAWKDYETKTKISEKEKEHAKLHGMIRTEITG	200
Qy	181	AEIAEEMEKERRFFQLOWCEVLLKAVNEIKKGVLDLQNLKYPHAQCNFPQDGLKAVES	240
Db	201	AEIAEEMEKERRFFQLOWCEVLLKAVNEIKKGVLDLQNLKYPHAQCNFPQDGLKAVES	260
Qy	241	LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLKLSALQVEQKE-----DS	285
Db	261	LKQYIEKLAADLYNIKQTQDEEKKQLTALRDLIKSSQLDLPKEVGLYVSRANSRRDS	320
Qy	286	QIROSTAYSLHQPOGNEKHGTERNGSLYKKSDDGIRKVMQKRCVXGNGFLTISHGTANRP	345
Db	321	QSROG-GYSMHQLQCNKEYGSEKKGFLKKSDDGIRKVMQKRCVXGNGFLTISHGTANRP	379
Qy	346	PAKLNLLTCOVKTNPKEKKCFDLISHDRTYHFAQDEQECQIWMVSLONKEEALNNAPK	405
Db	380	PAKLNLLTCOVKTNPKEKKCFDLISHDRTYHFAQDEQECQIWMVSLONKEEALNNAPK	439
Qy	406	GDDNTGNNIVQELTKETIISEVQRMGNDVCCDGPDPPTWLSTNLGTLTICESGIHRE	465
Db	440	GEQSTGNSL-EDLTKAIEDVQLPGNDICCDGSESEPTWLSTNLGTLTICESGIHRE	498
Qy	466	LGVHSPMQSLTLDVLTGTSSELLAKNIGNAGFNIMECCLPADSVKPNPGSDNNARKDY	525
Db	499	MGVHISIQSLDLKGLTSELLAKNIGNAGFNIMECCLPADSVKPNPGSDNNARKDY	557
Qy	526	ITAKYIERRYARKKHADNAAKHSLCEAVKTRDFGLLOAYADGVDTTEKIPLANGHEPD	585
Db	558	ITAKYVDRHFRSRKTCASSAKNELLEAIKQVYAGVELMEPL-LBSPQGLG	616
Qy	586	ETALHLAVRSVDRSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKA	645
Db	617	ETALHLAVRTADQTSLLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKA	676
Qy	646	SIETANESGETPLDIARKLKHCEELLTQALSGRFSNHVHEYEWRLHLLHDLDESDDDM	705

Db 677 TVDIVNQNETALDIKRLKATQCEDLLSQAKSGKFNPHVVEYNLRQDEMDESDDDL 736
Qy 706 DEKLQPSNREDDPISFYQLSGNQLQSNVSLARDAAANLAKEXORAFWPIILQNETYGA 765
Db 737 DDKSPFKKERSPPRQSFCHSSISPDQ---KLALPGFSTPRDKQRL---SYGA 784
Qy 766 L---LSGSPPPAQAAPSTTAPPLPPRNVGK---VQTASSANTLMKTNV 810
Db 785 FTNQIFASTDLPFSP-TSEAPLPPRNAGKPTGPPSTPLPLGTQTSSGSTLSKKRPP 843
Qy 811 SVDGSRQRSSDPPPAVHPPPLP----- 833
Db 844 PPPPG-HKRTLSDDPSPLPHGPNKGAIPWGNVDPSSSKTANKFGLSQOASTSSAKT 902
Qy 834 -----LRTSIN-----PLTPTPPPPV 850
Db 903 ALGPRVLKPLQKVALRKRTKTSHLSLDRTNIPPTFKSSQLTLPQKPLGELPPKPV 962
Qy 851 --AKTPSVMEALSQPSK-----PAPGISQIRPPP---LPP----- 881
Db 963 ELAPKPVGELPPKPGELPPKQGLDLPKQLSDLPKPKQMKOLPPKQGLDLAKSQ 1022
Qy 882 -----QPPRLPQKPAQGTDKSTPLTNKGQPRGVDLSATEALGPLSNMVLQPPAP 934
Db 1023 GDVSAKVQPPSEVTR-----SHTGDLSPNVQSRDAIQKQASEDSNDLTPTLP-ETEPV 1075
Qy 935 MPRSQATKLKPKRVKALYNCAVDNPDLTSEGDVVIIVGDEEDQEWIGHIDGPGK 994
Db 1076 LPRKINTGKNKVRVKTIDYQADNDDELTFIEGEVVIIVTGEEDQEWIGHIEGPERKG 1135
Qy 995 APFVSFVHFID 1006
Db 1136 VPPVSFVHILSD 1147
RESULT 8
DDF1_HUMAN
ID_DDF1_HUMAN STANDARD; PRT; 956 AA.
AC Q9ULH1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent Arf1 GTPase-
activating protein (PIP2-dependent Arf1 GAP) (ADP-ribosylation factor-
directed GTPase-activating protein 1) (ARF GTPase-activating protein
1) (Development and differentiation-enhancing factor 1) (fragment).
GN Name=DDEF1; Synonyms=ASAP1, KIAA1249;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Bone marrow;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Amid C., Osanger A.,
RA Pobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 11-959 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Chara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=99038209; PubMed=9819391;
RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,
RA Randazzo P.A.;
RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that

RT associates with and is phosphorylated by Src.";
RL Mol. Cell. Biol. 18:7038-7051(1998).
CC -!- FUNCTION: Possesses phosphatidylinositol 4,5-bisphosphate-dependent
CC GTPase-activating protein activity for ARF1 (ADP ribosylation
CC factor 1) and ARF5 and a lesser activity towards ARF6. May
CC coordinate membrane trafficking with cell growth or actin
CC cytoskeleton remodeling by binding to both SRC and PIP2. May
CC function as a signal transduction protein involved in the
CC differentiation of fibroblasts into adipocytes and possibly other
CC cell types (By similarity).
CC -!- ENZYME REGULATION: Activity stimulated by phosphatidylinositol
CC 4,5-bisphosphate (PIP2) (By similarity).
CC -!- SUBUNIT: Homodimer. Interacts with SRC and CRK (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-
CC associated; partially (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=2;
CC IsoId=Q9ULH1-1; Sequence=Displayed;
CC Name=1;
CC IsoId=Q9ULH1-2; Sequence=VSP_008365;
CC -!- DOMAIN: The PH domain most probably contributes to the
CC phosphoinositide-dependent regulation of ADP ribosylation factors
CC (By similarity).
CC -!- PTM: Phosphorylated on tyrosine residues by SRC (By similarity).
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -!- SIMILARITY: Contains 1 Arf-GAP domain.
CC -!- SIMILARITY: Contains 1 C4-type zinc finger.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; BX537768; CAD97831.1; -;
CC EMBL; AB033075; BAA66563.1; -;
CC IntAct; Q9ULH1; -;
CC Genew; HGNC:2720; DDEF1.
CC MIM; 605953; -;
CC InterPro; IPR002110; ANK.
CC InterPro; IPR001164; hRIP_like.
CC InterPro; IPR000108; Neu_Cyt_Fact_2.
CC InterPro; IPR001849; PH.
CC InterPro; IPR011036; PH related.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00023; Ank; 2.
CC Pfam; PF01412; ArfGap; 1.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF00018; SH3; 1.
CC PRINTS; PR00499; P67PHOX.
CC PRINTS; PR00405; REVINTRACTNG.
CC PRINTS; PR00452; SH3DOMAIN.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00248; ANK; 2.
CC SMART; SM00105; ArfGap; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS50088; ANK_REPEAT; 1.
CC PROSITE; PS50115; ARFGAP; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC PROSITE; PS50002; SH3; 1.
CC Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;
CC SH3 domain; zinc-finger.
CC NON_TER 1
FT DOMAIN 151 243 PH.
FT DOMAIN 266 389 Arf-GAP.
FT ZN_FING 281 304 C4-type.


```
FT REPEAT 427 459 ANK 1.
FT REPEAT 463 492 ANK 2.
FT DOMAIN 610 820 PRO-rich.
FT DOMAIN 894 956 SH3.
FT VARGPLIC 130 130 E -> ESSR (in isoform 1).
FT SEQUENCE 956 AA; 105967 MW; 68CB7034EC5B80 CRC64;

Query Match 46.8%; Score 2465.5; DB 1; Length 956;
Best Local Similarity 53.3%; Pred. No. 4.2e-96;
Matches 520; Conservative 110; Mismatches 204; Indels 141; Gaps 21;

QY 154 ETKITKEKEKEHEKAKHCHMIRTEISGAEIAEMEKERRFFOLQCEYLLKVNKIKKG 213
Db 1 ETKITKEKEKEHEKAKHCHMIRTEISGAEIAEMEKERRFFOLQCEYLLKVNKIKKG 60
QY 214 VDLLQNLKLYFHAQCNFFQDGLKAVESLKPSITETLSTDLHTIKQAQDEERRLOLRDIL 273
Db 61 VDLLQNLKLYFHAQCNFFQDGLKAVESLKPSITETLSTDLHTIKQAQDEERRLOLRDIL 120
QY 274 KSALQVEQKEDSQRISTAYSILHQPQGNKEHGTERNGSLYKSDGIRKVMQKRCVKNG 333
Db 121 KSSLLQDKEDSQSRQG-GYSMHQLQGNKEYGSEKGYLLKSDGIRKVMQKRCVKNG 179
QY 334 FLTISHGTANRPAPKLNLLTCQVKNPEKCFDLISHDRTHFOAEQBEQCIWMSVLQ 393
Db 180 ILTISHATSNRPAPKLNLLTCQVKNPEKCFDLISHDRTHFOAEQBEQCIWMSVLQ 239
QY 394 NSKEALANAFKDDNTGNNIVOLTEILSEVQRMGTNDVCCDGPADPTWLSNGLI 453
Db 240 NSKEALANAFKDDNTGNNIVOLTEILSEVQRMGTNDVCCDGPADPTWLSNGLI 298
QY 454 LTCIECSGHRHLGVHYSPMSLTLDVLGTSLLAKNIGNAGNEIMECCLPABSDSVKP 513
Db 299 LTCIECSGHRHMGVHISRIQSLDGLTSELLAKNIGNAGNEIMECCLPABSDSVKP 357
QY 514 NPGSDMNARKOYITAKYIERRYARKHADNAKHLCEAVKTRIDIFGLQAYADVDLT 573
Db 358 TPSSDMTVRKEYITAKYIDHRFPKSRCTSTSAKLNELLEAKSRDLALIQVYABGVLM 417
QY 574 EKIPLANGHEPDETALHLAVSDVTSLSHIVDFLVNCGNLDKOTKGSSTALHYCCLTDN 633
Db 418 EPL-LEPQOELGETALHLAVRTADQTSLSHIVDFLVNCGNLDKOTKGSSTALHYCCLTDN 476
QY 634 AECLKLLRGKASIBIANESGETPLDIAKRLKHECEBELLTOALSGRFNSHVHYEWRL 693
Db 477 PECLKLLRSKPTVDIVNQAGETALDIAKRLKATOCEDLLSQAQSGKFNPHVHYEWRL 536
QY 694 LHEDLDESDDMDKQLQSENRRDRPISFYOLGNSQLQSNVSLARDAANLAKEQRAF 753
Db 537 RQEEIDESDDLDLDDKPSPIKERSPRQSFCHSSISPDQ---KLALPGFSTPRDKQL- 592
QY 754 MPSILQNETYALLSG---SPPAPAPASTTSAPLPRNVGK-----VOTA 798
Db 593 ---SYGAFTNQIFVSTSDTSPTSP-TTEAPPLPRNAGKGTGPPSTPLSTQTS 643
QY 799 SSANTLWNTSVSDGGRSSRSDPPPAVHPLPLPLR----- 835
Db 644 SGSSTLSKRRPPPPPPG-KHRTLSDDPPPLPHGPNKGAVPWNGNDGSSSSKTKTKFEG 702
QY 836 -----VTSTNP-LTPTPPPVPA--KT-----PSVMEALSQ-----PSKPA 867
Db 703 LSQSSSTSSAKTALGPRVLPLKLPQKVALRKTDHLSLDKATIPPEIFQKSSQLAELPQKPP 762
QY 868 P-----PGISQIRPPP-----LPQPP-SRLPQK----- 890
Db 763 PGDLPPKPTLAPKFOIGDLPPKPGELPPKPOLGDLPPKPOLSLPPKPKMDLPPKPOL 822
QY 891 -----KPAPGTDKSTP--LTNKGQPRGPVDLSATEALGPLSNAMVLQP 931
Db 823 GDLLAKSQTGDSVPAQGPSEVTLKSHFLDLSFPNVQSRDAIQKQASESDNLTPTLP-ET 881
QY 932 PAPMPKSOATKLPKRVKALYNVADNPDELTFSEGDIIVDGBEDQEWIGHIDGDPG 991
```

```
Db 882 PVPLPRKINTGNKVRVKTIYDQADNDDLTETGESEVIIITGSEDOEMWIGHIEGQPE 941
QY 992 RKGAPPVSFVHFIAAD 1006
Db 942 RKGVPFVSFVHILSD 956

RESULT 9
Q8TDY4
ID Q8TDY4 PRELIMINARY; PRT; 903 AA.
AC Q8TDY4,
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Up-regulated in liver cancer 1.
GN Name=UPLC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Okabe H., Furukawa Y., Kato T., Hasegawa S., Nakamura Y.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 PH domain.
DR EMBL; AB056749; BAB85677.1; -.
DR HSSP; O7SIG6; 1DCQ.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002034; AIPM/HcIt_synth.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTNG.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; UNKNOWN_1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW ANK repeat.
SQ SEQUENCE 903 AA; 99155 MW; D5E89518AFC86656 CRC64;

Query Match 39.3%; Score 2071.5; DB 2; Length 903;
Best Local Similarity 43.4%; Pred. No. 1.6e-79;
Matches 446; Conservative 164; Mismatches 271; Indels 147; Gaps 21;

QY 1 MPDQISYSEFVAETHEDYKAPT-ASSFTTRTAQCRNTVAALIEEALDVRMVLKMKKSVK 59
Db 1 MPEQFSAEFLAVTAEDLSSPAGAAFAAKMPYRGAALAREEILEGQAILQRIKAVR 60
QY 60 AINSGSLAHVNEBOYTOALEKFGNCVCRRDDPDLSGAFKPSVFTKELTALFKQLIQNM 119
Db 61 AIHSSGLGHVNEBOYREAVESLGNLSHLSQNSHELSTGLNLAVFTREVAALFKQLIQNL 120
QY 120 NNIIISFPLDLSLKGDLKGVKGDLPKPPDKAWDYETKITTKIEKEKHAHLGHMIRTEIS 179
Db 121 NNISFPLDLSLKGDLKGVKGDLPKPPDKAWDYETKITTKIEKEKHAHLGHMIRTEIS 174
QY 180 GAETAEEMEKERRFFOLQCEYLLKVNKIKKGVDLQNLNLIKVFHAQCNFQDGLKAVE 239
Db 175 PGEVAQDMQREERIFQLHMCYELLKAGESQMKQGPDLFQSLIKVFHAQCNFQDGLKAVE 234
QY 240 SLKPSIETSLDHTIKQAQDEERRLOLRDILKSLAQVEQKEDSQRIST--AYSJHQ 297
```

```
Db 235 SLFPFIEKLAASVHALHOAQDELOKLTQLRDSLRGTLQLESREBHLSRKNSGCCYSIHQ 294
Qy 298 PQGNKEHGTGNSLYKSDGIRKVKQKSCVKNGLTISHGTANRPPAKNLITTCOVK 357
Db 295 HQGNKQFTEKVGFLYKSDGLRRVWQKKGCVKGYCLTISHSTINRPFVKLTLLTCQVR 354
Qy 358 TNPBEKKCFDLISHDRTYHFOAEDRQEQIWMKSVLQNSKEBALNNAFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTYHFOAEDHECEAWVSVLQNSKDEALSAFGEBSGPGSGWS 414
Qy 410 TGENNIVOLFKELIIESEVORMTNDVCCGAPDPTWLTSTNLGILTCIECGIHRELGVH 469
Db 415 AGHDGEPDLTKLJIAEVKSRPNSQCCDCAADPTWLTSTNLGVLTCIQCSGVHRELGV 474
Qy 470 YSPQSLDVLGTSSELLAKNAGNEIMECCLPRAEDSVKPNPGSDMNAKDYITAK 529
Db 475 FSRMQSLDLDLGPSELLALNMGNSTFNEVMEALPQSHGPKPSAESDMGTRRDYINAK 534
Qy 530 YTIERYARKKIADNAAKLHSLCEAVKTRDIFGLIQAYADGVDLTEKIPLANGHEPDETAL 589
Db 535 YVEHRFARR-----CTPEQRQLWTAICNRDLSLVLEAFANGQDFGQPLPGPAQAPEELVL 590
Qy 590 HLAIVRSVDRTSIHIVDFLVQNSGNDLKQTKGKSTALHYCCCLTDNAECLKLLRKGASIEI 649
Db 591 HLAIVKVAQASLPLVDFIIQNGCHLDAKADGNTALHYAALYNQPCDCLKLLKGRALVGT 650
Qy 650 ANESGETPLDIAKRLKHECHCELLTOALSGRFNSHVHYEWRLHEDLDESDDMDKEL 709
Db 651 VNEAGETALDIARKKHKECELELQQAQGTFAFPLHVDYGVNISTEPGSDSEDEEBK- 709
Qy 710 QPSENREDRPTSFYQLGNSQLQNSVSLAADAAMLAKEKORAFMPSILQNETYGALLS- 768
Db 710 -----RCLLKLPQAQWASGRLLDISNKTYETVASL 739
Qy 769 GSPPPAQAAPSTTSAPLPPRNKGVQTASANTLWKTNSVSDGGRQSSSDPPPAVH 828
Db 740 GA-----ATQGESEDCPPLPVKN-----SSRTLQV-----GCARHASGDRSEVS 780
Qy 829 PPLPLRVTSINPLTPTPPPPVAKTPSVMEALSQSKP--APPGISQ--IRPPPLPPQ-- 883
Db 781 -----SUSSEAPETPESLGPSPASSLSPL-EPGDFSQAPPNSEGLEPPGTSRPSL 833
Qy 884 -----PSRLPOKKPAPGTDKSTPLTNKGQPRGPFVDSLATEALGPLSNAMVLQPPAPMPRK 938
Db 834 TSGITPSEM--YLPVRFSESTRYRG-ANSPED-----GPSAR-----QPLPRR 876
Qy 939 SQATKLKPKRVKALYNVCVADNPDELTFSEGDVVIIVDGEEDQEWMTGHIDGDPGRKAPV 998
Db 877 NVP-----VGIPEGDGSRTGSLPA 895
Qy 999 SFVHFIAD 1006
Db 896 SSVQLLQD 903

RESULT 10
Q6P9F4
ID Q6P9F4 PRELIMINARY; PRT; 903 AA.
AC Q6P9F4;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Up-regulated in liver cancer 1.
GN Name=UPLC1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

```
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Abramo R.D., Mullaly S.J.,
RA Raha S.S., Loquellano N.A., Peters G.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 PH domain.
DR EMBL; BC060786; AAH60786.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002034; ALPM/Hcit_synth.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR Pfam; PF00023; Ank; 1.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTNG.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS00815; ALPM_HOMOCIT_SYNTH_1; UNKNOWN_1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP_1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR ANK repeat.
KW ANK repeat.
SQ SEQUENCE 903 AA; 99126 MW; D291DF00FD82797D CRC64;

Query Match 39.2%; Score 2067.5; DB 2; Length 903;
Best Local Similarity 43.3%; Pred. No. 2.4e-79;
Matches 445; Conservative 164; Mismatches 272; Indels 147; Gaps 21;

Qy 1 MPDQISVSEFVAETHEDYKAPT-ASSFTTRTAQCRNTVAATEALDVRMVLVYKMKGVK 59
Db 1 MPEQFSVAEFLVTAEDLSSPAGAAFAAKMPRYGAALAREEILLEGDAQILRIKKA 60
Qy 60 AINSSGLAHVENEEOYTQALEKFCGNCVCRRDDPDLSAFLKFSVFTKLTALFKNLIONM 119
Db 61 AIHSSGLGHVENEEOYREAVESLGNLSHLSTGFLNLAVFTREVALFKNLIONL 120
Qy 120 NNIISFPLDLSLLKGDGLKGVKDKLKKPPDKAWDKVETKITKEKKEKHAHLGMRITBIS 179
Db 121 NNIVSFPLDSLMLKGLRGRQDSKKQLEKAWDKVEAKWAKLEKE-RDRARVTGGI----- 174
Qy 180 GAETAEEMEKERRFPQLQCMCEYLLKVNIEKITKGVVDLLQNLKIYFHAQCNFFQDGLKAVE 239
Db 175 PGEVAQDMQERRIFQLHMCVEYLLKAGSQMKQPGDFLQSLIKFFHAQHNFQDQWKAQ 234
Qy 240 SLKPSIETLSTDLHTIKQAQDEERRQLQLDLILKLSALQVEKEDSQSQST--AYSILHQ 297
Db 235 SLFPFIEKLAASVHALHOAQDELOKLTQLRDSLRGTLQLESREBHLSRKNSGCCYSIHQ 294
Qy 298 PQGNKEHGTGNSLYKSDGIRKVKQKSCVKNGLTISHGTANRPPAKNLITTCOVK 357
```

```

Db 295 HGNKQFGTEKGYFKYKSDGIRRAWQRKCKGVKGLTISHSTINRPPVKLTLLTCTQVR 354
QY 358 TNPEKKCFDLISHDRTYHFQADEQEQIWMVSLQNSKEEALNNAFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTYHFQADEHECEAWVSLQNSKDEALSAFLGEPGAGSGS 414
QY 410 TGNNIVVELTKEIISVQRTMGNDVCCDGAADPTWLTSTNLGILTCIECSGIRHRELGVH 469
Db 415 AGHDGEPHDLTKLLAEYKSPGNSQCCDGAADPTWLTSTNLGILTCIECSGIRHRELGVH 474
QY 470 YSPMOSLTDLVLTGSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMNARKDYITAK 529
Db 475 FSRMOSLTDLVLTGSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMNARKDYITAK 534
QY 530 YIERYARKHADNAKLSHCEAVKTRDIFGLQAYADGVDLTETKIPLANGHEPDETAL 589
Db 535 YVEHRFARR---CTPEQRQLWTALCNRDLLSLEAFANGQDFGQPLPGPDAQAEELVL 590
QY 590 HLAHSVDRSTLSHIVDFLVQNSGNDKOTGKSTALHYCCLTDNACELKLLRGKASIEI 649
Db 591 HLAHVANQASLPLVDFIIONGGHLDAAKADGNTALHYAALYNQPDCLKLLKGRALVGT 650
QY 650 ANESGETPLDIAKRLKHECELLTQALSGRPNSHVVEYEWRLHEDLDESDDMDKEL 709
Db 651 VNEAGETALDIARKKHKECEBELLEQAQAGTAPFLPHVDYISWVISTEPGSDSEDEBEK- 709
QY 710 QPSENREDRPISFYQLGNSQLQNSAVSLARDAANLAKEKORAFMPSILQNETYVGLLS- 768
Db 710 -----RCLLKLPAQAHWASGRDLISNKTYETVASL 739
QY 769 GSPPPAQPAASTTSAPPLPRNVKGVQTSANTLWKTNSVSDVGGSRQSSDPAPVH 828
Db 740 GA---ATPQGESEDCPPPLPVKN-----SSRTLQ-----GCARUASGDRSEVS 780
QY 829 PPLPLRVTSINPLTPPPPVAKTPSVMEALSQSKP--APPGISQ-IRPPPLPPQ-- 883
Db 781 -----SUSSEAPETPESIGSPASSSLMSPL-EPGDPSPAPNSEEGLREPPGTSRPSL 833
QY 884 -----PSRLPQKPPAGTDKSTPLTNKQPRGPVDSLATEALGPLSNAMVLQPPAPMPRK 938
Db 834 TSGITPSEM--YLPVPFSSESTRVRG-ARSPED-----GPSAR-----QPLPRR 876
QY 939 SOATKLKPKRVKALYNCVADNPDELTFSEGDIIVDGEDQEWIIGHIDGDPGRKAPPV 998
Db 877 NVP-----VITEGDSRGTGSLPA 895
QY 999 SFVHFAD 1006
Db 896 SSVQLQD 903

```

RESULT 11
Q7QJY3

```

ID Q7QJY3 PRELIMINARY; PRT; 1015 AA.
AC Q7QJY3;
DT 01-MAR-2004 (TREMELrel. 26, Created)
DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE AGCP3317.
GN Name=agCG57483; ORFNames=ENSGANGG00000018249;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

```

```

CC Preliminary data.
DR EMBL; AAB01008807; EAA04701.1; -.
DR HSSP; Q60631; IGBQ.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR001452; SH3.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3.1; 1.
DR PRINTS; PR00405; REVINTRACTNG.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR PROSITE; PS00115; ARFGAP; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 1015 AA; 111957 MW; DE8972F0E6CC898C CRC64;

Query Match 28.4%; Score 1499.5; DB 2; Length 1015;
Best Local Similarity 32.5%; Pred. No. 2.5e-55;
Matches 387; Conservative 150; Mismatches 294; Indels 359; Gaps 25;

QY 1 MPDOISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMLVKMKSKVA 60
Db 1 MPELIGVEFEETREDYNSPTSTFVSRMAQCROTIAETLEETLDFDREGLTCLKKAKVA 60
QY 61 INSSGLAHVNEEQYQTALEKFGNCVCRDDPDLGSAFKFSVFTKELTALFNLIQNMM 120
Db 61 IHNSGNTHVDNEMCLVRALERLGSVALSKEEPPDIGAAFLKFSVVTKELLSALMKTLMQIN 120
QY 121 NIISFPDLSLKGDLGVKGLKPPFKAMKDYETKITIEKKEKHAHLGHMITEISG 180
Db 121 NIVMFPVDSLKSLRGKMGEMKPPDKAAKDYDSKLMKIEKKEKALAKEVGMRTVTP 180
QY 181 AEIAEEMEKRRFPOLQMCVLLKNEIKIKGVVDLQNLIKYFHAQCNFPDGLKAVES 240
Db 181 AEIAEIEKERRVPOLQMCVLLKNEIKIKGVVDLQNLIKYFHAQCNFPDGLKAVES 240
QY 241 LKPSIETLSTDLHTIKQAQDEERQLIQLRDILKASAL---OVEQKEDSIQIRSTAYS LHQ 297
Db 241 FGTYIEELS IKQIRHKQDEERKLELRLLASTPDPDRVENVPSCDKSAGIYSLHQ 300
QY 298 PQGNKEHGTENGSLYKKSDDGIRKVMQKRCSVKNQGLTISHGTANRPAPKLNLLTQVK 357
Db 301 LQGDKNHD----- 308
QY 358 TNPEKKCFDLISHDRTYHFQADEQEQIWMVSLQNSKEEALNNAFKGDNDTGENNIVQ 417
Db 309 -----NRSYHFOADEADQKAWMSVLINCKEKALAKAFQHANPQMSPLI- 353
QY 418 ELTKETIIEVQRTMGNDVCCDGA-PDPTWLTSTNLGILTCIECSGIRHRELGVHSPMOSL 476
Db 354 ELQTVIKHILQNLPGNDCCDGSNDVTWLSLNFGLVLCIQC SGVHRLDGLVHHSRIQSL 413
QY 477 TLDVLGTSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMNARKDYITAKYIERYA 536
Db 414 TLDNLTTAQLLVARAMGNALNEVMEATLA---QSKLTPESTMEERYDFIRAKYVAKYV 470
QY 537 RKHADNNAKLSHCEAVKTRDIFGLQAYADGVDLTETKIPLANGHEPDETALHLAVRSV 596
Db 471 MRTCADDRDLRNDLEQAVINADLQQLQVWAEAGDLTCVLP----- 511
QY 597 DRTSLHIVDFLVQNSGNDKOTGKSTALHYCCLTDNACELKLLRGKASIEIANESGET 656
Db 512 ----- 523
QY 657 PLDIAXRLKHECEBELLTQALSGRPNSHVVEYEWRLHED- LDESDDD---NDEKLQ 711
Db 524 AFD-----HINTDWN-IHDDGSTDFSDDDTALMDER--- 553
QY 712 SENREDRPISF-----YQLSGNQLQNSAVSLARDAANL-----AKE 748
Db 554 ---KSRSRPPFVGSDSPVALRSRSTCDISQSGSSPSSSCNPNIARDQRIPIPPSSGT 610

```

```
QY 749 KQAFMPSILQNETYCALLS-----GSPPPAQPAAPTSTTSAPPLPRNVKQVQTA 798
Db 611 SPKQNTSTVFSHVASGTSSVQNSIGMIGVSAFTSGSSPIGNSSSVQGVSTSOIT 670
QY 799 SSSA-----NTLWKTNSVSDGSRQRSSDPPAV----- 827
Db 671 SSSVATFNTSTNKKFSSVNV-GSLKKRTAPPPPTTYGTLPAPRHSQNLDELFGVSSQ 729
QY 828 -----HP-----PLPPLRVTSNPLTPPTPPPVAKTPSVMEALSQPSKPAP----- 868
Db 730 HVLHGLHHHPDMYSTLPHLRGSDPISNNSASSVPAALLSSMRGGSTKSMQAQLLFDNK 789
QY 869 -----PGISQIRPP--PLPQPSPRLPQKPGAPGTDKSTPLTN--KGQ 907
Db 790 IPDREYPTSSGAASVAGAKLVLPAGEIPLQKPVSNRPKQIPQISIAVTQISNTEKSF 849
QY 908 PRGPVLDLSATEALGPLSNAMVLQPPAPMPKRSQA-----TKLKP----- 946
Db 850 ANG-----QSNESLTSVDDALIMRRKIKGPASSAGYDSNLDGDTKETSYSNYDTSGIRNA 905
QY 947 ----- 947
Db 906 LDNSVASSGGHMSLNSYSRNDNSGDASCVDLDSFNLSSRSGFGAMRRCALYDCNAD 965
QY 959 NPDELTTFSEGDVLIIDGGE--EDQEWIIGHIDGDPGRKGAFVSVFVHTIAD 1006
Db 966 NDELEFKEGEILIVINERTDDENWMEQIEGDSMRGMFPVSVFQMDD 1015

RESULT 12
Q6AWJ6 PRELIMINARY; PRT; 935 AA.
AC Q6AWJ6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE LP17217p (Fragment).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; BT015252; AAT94481.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IBA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3 2.
DR Pfam; PF00023; ANK; 2.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3_1; 1.
DR Pfam; PF07653; SH3_2; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTING.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
```

```
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; SH3 domain.
FT NON_TER 1
SQ SEQUENCE 935 AA; 102696 MW; 45BF94ED72BF0FC3 CRC64;

Query Match 26.2%; Score 1380; DB 2; Length 935;
Best Local Similarity 37.6%; Pred. No. 2.5e-50;
Matches 362; Conservative 127; Mismatches 299; Indels 174; Gaps 31;

QY 192 RFPOLQCEYLLKVEIKKGVLLQNLQIKYFHAQCFFQDGLKAVESLXPSIETLSTD 251
Db 1 RLYQLQCEYLLKYKDKTKTGIELLQHLIEYHALSNYFKQGLQTIHFQTYICDLSK 60
QY 252 LHTIKQAQDEERRQLIQLRDILKSALQVEQKED---SQIRQSTA-YSLHQPGQKNEHTE 307
Db 61 LHEIKQKQDEDRRSLLDLRTVLRSTPDFERVDNVPSSESRSRGAGYSLHQIQGDKHGV 120
QY 308 RNSLYKKSQG-IRKQVQKRKCSV-KNGFLTIHSHTANRPPAKNLLTCVKTNPKEKCC 365
Db 121 RQGHLLKKSEKGRVRVQKRCRVTSDFLDFIHAEDESKPPTRVNLTLTCQIKPVPDDKRG 180
QY 366 FDLISHDKTYHFAEDEQECQIMNSVLQNSKEEALNNAFKGDNDTGENNIVQELTKEIS 425
Db 181 FDLISYRPHYHFAEDEGDQKAWAVLVNCKEALTKAFQHANPQMSPLV-ELQKTVIR 239
QY 426 EVQRTGNDVCCDCA-PDPTWLSTNLGILTCIECSGIHRELGVHYHSPQSLTLDVLGTS 484
Db 240 YVQLLPNGDRCCDGRNDVTWISLNFILVCIQCSGVHRDLGVHHSRIQSLTLDNLTTA 299
QY 485 ELLAKNIGNAGFNEIMECLPAEDSVKPNPGSDMNARKOYITAKYIERRYARKKHADNA 544
Db 300 NLLTARANGNSTLNDIMEAKL---GRGKLOHESMEERYDFIRAKYAKRVYMTCTSDDN 356
QY 545 AKLSLCEAVKTRDIFGLLQAYADGVDLTKIPLANGHEPDEPTALHLAVRSVDRSTLHV 604
Db 357 DLRCLEQAVNVADMSQLQVWAGADLTCLP---SSDAGETALHLAVLREMSTLHV 413
QY 605 DFLVQN-----SGNLDKQTKGSGTALHYCCLTDNAECLKLLLRKASIEIANE 652
Db 414 DFLIQNPMPKGLNKATNPAGLLD-VTGGK-NTALHLCALHRRRCWKLLLRSGADYELKNS 471
QY 653 SGETPLDIKRLKHEKCEELTQALSGRFNSHVHVEYEWRLLED--LDESDDD--WDEK 708
Db 472 QNKALDIAKEMGHNSCRELIECAIKREKSAFHDINTDNLNPNEDSGTDFSDDETVIDER 531
QY 709 LQPS-----ENRREDRP-----ISFYQ-LGS-----N 729
Db 532 SSSSPIANCPSRQFTLPGLPSYTHSAGTSPKQHSVQYLGSAITNVGNGPNGSGSSPS 591
QY 730 QLQSNVSLARDAANLAKEKORAFMPSITLQNETYCALLS-----PPPAQ 775
Db 592 SASQSQVRAARNSLN-----MQSDLGHVGTGARKSTSTANWNSLKRTAPAPPPGT 642
QY 776 PAAPSTTS---APPLPRNVKQVQTASSANTLMKTNVSU-----DGG 815
Db 643 LGSASSSFFYGTLPHPPRHSQNFDAIDIRAINHNKQSLDVAYGTLPHLRSVSSPRGGGG 702
QY 816 SRQRSSSDP---PAVHPPLPLRVTSNPLTPPTP-----PPVAKTPSV 856
Db 703 YGYGVSDPQGGSGNGSNNSLNPAMTTFCHKKSPPGESLNRNIHLAGAKLVLPPTGELPTL 762
QY 857 ----MEALSQPSKAPPAGISQIRPPPL-----PPQPSRLPQKKPA 893
Db 763 KHVDSKALTREP-KIP-PPG-----PPSREISNGQSNESSISMDDEGPVAPPRKLVNQ-S-A 814
QY 894 PGTDKSTPLTNKGQRPVLDLSATEALGPLSNAMVLQPPAPMPKRSQATKLKPK----- 947
Db 815 NFPDYESWHTDMSSGGGLDHSASNSVSSSDNDRLNSSP-DNPSKTKGAGLGKGFHYNGQ 873
QY 948 -RVKALYNCVADNPDELTFSEGDVLIIDGGE--EDQEWIIGHIDGDPGRKGAFVSVFVHTI 1004
```

```
Db 874 RRCALYDCVADNDELEKEGEILVUNERTDDENMMEGIEGQPTRKGMFPVSVFVHML 933
Qy 1005 AD 1006
Db 934 PD 935

RESULT 13
Q8N282
ID Q8N282 PRELIMINARY; PRT; 208 AA.
AC Q8N282;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ33802.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tongue;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami J., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hibi T., Kobatake N., Inagaki H., Ikemura Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
CC -!- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AK091121; BAC03588.1; --
DR HSSP; Q9N2M3; 1UFF.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR Pfam; PF00018; SH3_1; 1.
DR PRINTS; PR00499; p67PHOX.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 208 AA; 21869 MW; 8ABA541403C99B6F CRC64;
```

Query Match 20.4%; Score 1075.5; DB 2; Length 208;
Best Local Similarity 81.4%; Pred. No. 2.8e-38;
Matches 206; Conservative 1; Mismatches 1; Indels 45; Gaps 1;

```
Qy 754 MPSILQNETYCALLSGSPPPAQAAPSTTSAPPLPPRNVGVQVOTASSANTLWKTNSVSD 813
Db 1 MPSILQNETYCALLSGSPPPAQAAPSTTSAPPLPPRNVGVQVOTASSANTLWKTNSVSD 41
Qy 814 GGSQRSSDDPPAVHPPPLPLRVSTNPLTPTPPPPVAKTFSVMEALSQSPKPPPGISQ 873
Db 42 -----DPLTFTPPPPVAKTFSVMEALSQSPKPPPGISQ 75
Qy 874 IRPPPLPPQPSRLPQKKPAPGTDKSTPLTNKGQPRGPDLSATEALGPLSNAMVLQPPA 933
Db 76 IRPPPLPPQPSRLPQKKPAPGADKSTPLTNKGQPRGPDLSATEALGPLSNAMVLQPPA 135
Qy 934 PMPRKSQATKLKPRVRYALYNCVADNPDELTFSGDVIIVDGEEDQEWIIGHIDGDPGRK 993
Db 136 PMPRKSQATKLKPRVRYALYNCVADNPDELTFSGDVIIVDGEEDQEWIIGHIDGDPGRK 195
Qy 994 GAPPVSVFHFAD 1006
Db 196 GAPPVSVFHFAD 208

RESULT 14
Q6DEH1
ID Q6DEH1 PRELIMINARY; PRT; 307 AA.
AC Q6DEH1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler N.K.,
RA Altschul S.P., Zeeberg B., Buotow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalley D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Genes and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Singapore local strain; TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC077145; AAH77145.1; --
KW Hypothetical protein.
FT NON TER 307 307
SQ SEQUENCE 307 AA; 35463 MW; 5BE7CD31B38F5FDC CRC64;
```

Query Match 16.7%; Score 878; DB 2; Length 307;
Best Local Similarity 54.8%; Pred. No. 9.6e-30;
Matches 167; Conservative 70; Mismatches 64; Indels 4; Gaps 2;

Qy 1 MPDQISSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMLVKMKSVKA 60

```

1  MPEKISIDFVSLAKEDLSSPGSGFQSKMGDCRSTVAALBESLEQDQMSLQKMKKIYKA 60
61  INSSGLAHVNEEQYTOALEFGGNCVCRDDPDLGSAFLKFSVTKELTALFKNLIQNNV 120
61  IHNSGLAHVNEKEQYTEVLNGLSHLSQDNNEVSTGFLNLAFTREVYALFKNLVQNLN 120
121  NIISFPLDLSLLKGLDKGVKGLDKKPFDKAWDKYTKIKIEKEKEHAHKLHGMTRTETSG 180
121  NIISFPLENVLUKSELRDSRLKQMEKSNKYDAKIKLKERREKQKGLRLDWT- 179
181  AEIAEEMEKERRFFQLOMCEYLLKVKNEIKKGVDDLLQNLIKYFHAQCNFFQDGLKAVES 240
180  -DTAEDMERERNFQLOMCEYLLKVKNEIKKGVDDLLQNLIKYFHAQCNFFQDGLKAAEN 238
241  LKPIETLSTDLHIKQADERRLOLIRILKSALQVEKDSQIRQST--AYSLLHQP 298
239  LGPTFVEKLAASVHTVRHEQDEVKLSQLRSLRQLQVDGKEDFLNRKNSGHGYSIHQP 298
299  QGNKE 303
299  QKKKK 303

RESULT 15
Q9V367 PRELIMINARY; PRT; 762 AA.
AC Q9V367;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE CG30372-PA.
GN ORNames=CG30372;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.B.,
RA Abril J.F., Aghayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Godek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2];
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celnik S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3];
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnik S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4];
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betscounot B.R., Celnik S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5];
RP SEQUENCE FROM N.A.
RX SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6];
RP SEQUENCE FROM N.A.
RX SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AE003838; AAF59133.3; -.
DR HSSP; Q60631; 1GBQ
DR FlyBase; FBgn0050372; CG30372.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00018; SH3_1; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTING.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00088; ANK REPEAT; 1.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50002; SH3; 1.
DR ANK repeat; SH3 domain.
KW ANK repeat; SH3 domain.
SQ SEQUENCE 762 AA; 82178 MW; 9F0236BA1DA9B7FB CRC64;

```

Query Match

16.0%; Score 844; DB 2; Length 762;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 08:28:06 ; Search time 10576 Seconds
(without alignments)
4609.110 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYK.....DGDPRKGAFVSVFHFAD 1006

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09914042/runat_04082005_090632_2111/app.query.fasta_1.1159
-DB=genEmbl -QWRT=fastcap -SUFFIX=rgc -MINMATCH=0.1 -LOOPFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914042 @CGN 1 1 6279 @runat_04082005_090632_2111 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*

2: gb.hcg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sw.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5253	99.6	5711	9 AB007860	Homo sapi
2	4968.5	94.2	5544	9 BC063308	Homo sapi
3	4727	89.6	4713	10 AK172944	Mus muscu
4	4215.5	79.9	3989	5 BC070750	Xenopus l

ALIGNMENTS

RESULT 1	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	mRNA	linear	PRI 10-JAN-2004
LOCUS	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	mRNA	linear	PRI 10-JAN-2004
DEFINITION	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	mRNA	linear	PRI 10-JAN-2004
ACCESSION	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	mRNA	linear	PRI 10-JAN-2004
VERSION	AB007860.1	GI:2662080	5711 bp	mRNA	linear	PRI 10-JAN-2004
KEYWORDS						
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE						
AUTHORS	Ishikawa, K., Nagase, T., Nakajima, D., Seki, N., Ohira, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O.					
TITLE	Prediction of the coding sequences of unidentified human genes. VIII. 78 new cDNA clones from brain which code for large proteins in vitro					
JOURNAL	DNA Res. 4 (5), 307-313 (1997)					
MEDLINE	98116555					
PubMed	9455477					
REFERENCE						
AUTHORS	2 (bases 1 to 5711)					
TITLE	Direct Submission					
JOURNAL	Submitted (06-OCT-1997) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, Yata 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)					
FEATURES						
	Location/Qualifiers					

BC080847 Mus muscu
AR243296 Sequence
AR243295 Sequence
I74317 Sequence 15
I74314 Sequence 12
AR243298 Sequence
AR243297 Sequence
BC048818 Mus muscu
AF112886 Bos tauru
AR243292 Sequence
AF075462 Mus muscu
AF075461 Mus muscu
AR243294 Sequence
AR243293 Sequence
CQ720712 Sequence
AK122477 Mus muscu
BX537768 Homo sapi
AB033075 Homo sapi
CQ716704 Sequence
AX747456 Sequence
AK092291 Homo sapi
AX741097 Sequence
AB056749 Homo sapi
BC060786 Homo sapi
CR387820 Gallus ga
CQ575946 Sequence
CQ575945 Sequence
AC093104 Drosophi
AE003838 Drosophi
AC020344 Drosophi
AC008370 Drosophi
BT015252 Drosophi
AX746789 Sequence
AK091121 Homo sapi
CR385625 Gallus ga
AK000253 Homo sapi
BC077145 Dario rer
U92478 Mus musculu
BX950804 Gallus ga
AX361503 Sequence

source

```

1. 5711
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="HG1091"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1. 5711
/gene="KIAA0400"
<293...3361
/gene="KIAA0400"
/note="Start codon is not identified"
/codon_start=1
/protein_id="BAA23696.2"
/db_xref="GI:40788243"

```

gene

CDS

```

/translation="GGGAPAAVQRPRAEAMPDQISVSEFVAETHEDYKAPTASSFTT
RTACRNTVAATIEALQVDVRLVYKMKSVKAINSSGLAHVNEBEQYTALEKFGNCC
VRDDPDLSGAPLKFPSVFTKELTALFKNLIQNMNIIISPLDLSLLKGLKGVKGLKK
PFDKAWKDYETKIKIEKEKHAHLGMIETESGAEIAEMEKERRFPQLQWCEVL
LKVNEIKIKGVLDLQNLIKYFHAQCNFFQDGLKAVESLKPSIETLSLDLHIIKQAD
BERQLQRLDLSALQVEQKDSQIRQSTAYSLHQPQGNKEHGTERNGLYKKSOG
IRKWKQKRSKVGKFLTISHGTANRPAPKLLTLCQVKTNPKEKCKFDLSHDTYH
FQADEQPCQIWMVLSQNSKEBALNNAFKGDDNTGNNIIVQELTKEIIESEVQRTGND
VCCDGAPODPTWLSNLTGILTCIECGIHFRELGVHYSRMQSLTLDVLGTSLLAKNI
CAAFNEIMECLPAEDSVKPNQSDMMNARKDYITAKYIERVARKHADNAKLHSL
QVGNKTRIDFGLQAYADGVDLTEKIPLANGHEDEDTALHLAVRSVDRTSLSHIDFLV
CNSKRLDQKQSTALHYCCLTQNAECLKLLRLKASIEIANESGETPLDIKRLKH
EHCSELTLQALSFRNSHVHYEWRLLEHDESDDDMDKQLQFSPNRRDRPISFY
QLGNSQLQNSAVSLRADNALEKQRAFMSIILQNETYGALLSGSPPAQPAAPSTT
SAPPLPPNNGVKQVOTASSANTLWKTNSVSDGGRSRRSSDPAPVHPPLPLPLVTSTN
PLTPTPPNVAKTSVMEALSQPSKAPPGIISOIRPPPLPQPPSRLPKQKPPAGADK
STPLTNKQPPRGVDLSATEALGPLSNAMVLQPPAPMPKRSQATLKPKRVLKYNVC
ADNPDELTFSSGSDVIIVDGEEDQEWIIGHIDGDPGRKGAFFVPSFVHFITAD"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 3e-187 Length: 5711
Score: 253.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 9 Gaps: 0

```

US-09-914-042-1 (1-1006) x AB007860 (1-5711)

Qy	141	AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysLysLeuThrLysLysLeu	160
Db	761	GATCTGAAAAGACCTTTTGTATAAAGCTTTGGAGGACCTATGAAACAAAATAAACCAAGATA	820
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly	180
Db	821	GAAGAAGGAGAAAAGGAAACAGCCCAAGCTCCATCGGATGATTCGAGCTGAATAAAGCGGA	880
Qy	181	AlaGluLysLysGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu	200
Db	881	CGCGAAATTCGGAAGAGATGGAAGAGAGAGAGCGCTTCTTCCAGCTACAGATGTGCGAG	940
Qy	201	TyrLeuLeuLysValAsnGluLysLysLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	941	TATCTGCTGAAGGTCACAGAAATCAAGATTAAGAAGGAGTAGATTTTACTTTCAGAACTG	1000
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	1001	ATCAATACITTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAAGCCGTGGAAGC	1060
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1061	CTCAACCTTCCATTGAACCGCTGCTACGGATCTTCCACGATCTTCAACAGCAAGCCCGGAT	1120
Qy	261	GluGluArgArgGlnLeuLysGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	1121	GAAGAAAGAGGCGAGTGTATACAGCTTCGAGATATTTTGAATCCGATTCGAGGTGAA	1180
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1181	CAGAAAGGAGACTCCCAATTCGTTCAGAGCACAGCTTATAGCTTACATCAGCTCAGGGA	1240
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1241	AAACAGAAACATGGGACCGAGCGGAAACGGAGCCCTCTACAAGAAAGAGTGACGGGATCGA	1300
Qy	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1301	AAAGTGTGGCAGAAAAGGAAATGTTTCAGTTAAATGGTTTTCTGACCATATCCATGGT	1360
Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1361	ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGCTAACCTGCGCAGGTGAAGACCAACCT	1420
Qy	361	GluGlnLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1421	GAGGAGAAGAGAGTGTCTTACCTTATTTTCATGACAGAACTTACCACCTTTTCAAGCTGAA	1480
Qy	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu	400
Db	1481	GATGAACAGGAATGTCAAAATATGGATGTCGTGCTGCMAAATAGCAAGAAGAGCTTTA	1540
Qy	401	AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1541	AACATGCAITTAAGGGGAGTACAACTACGAGAAATAACATCGTCCCAAGAACTGACA	1600
Qy	421	LysGluLysLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1601	AAGGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGTCTGCTGTGACTGTGG	1660
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysValLeuGluCysSer	460
Db	1661	CGCCGAGATCTCATGCTTTCCACCAACCTGGGCATCTCTGACCTGCATCGATGTGTTCC	1720
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
Db	1721	GGATCCACCGAGAGCTGGGGGTTTCATTTATTCAGGATGACAGTCCCTGACCTTAGATGTA	1780
Qy	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1781	CTGGAAACATCTGAGCTGCTGCTGCCAAGAAATATTGGGAATGCGAGCTTTAATGAGATC	1840

QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
 Db 1841 ATGGAATGTGCTTACCAAGCTGAGACTCAGTCAAAACCCCAAGCAGGACATGAAT 1900
 QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
 Db 1901 GCAAGAAAGGACTACATCACAGCCAAAGTACATCGAGAGAGATACGCAAGAGAGAGAC 1960
 QY 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
 Db 1961 GCGGATAACGGCGGCGAAGCTTCACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTTT 2020
 QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 2021 GGATTGCTCCCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAATCCCACTGGCCAAAC 2080
 QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
 Db 2081 GGACATGAGCCGGATGAACCGCCCTCCACTTGCAGTTCAGATCCGTGGATCGAACCTCT 2140
 QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
 Db 2141 CTTCACTATTAGACTTTTATGTTAGTTCAGAAACAGTGGGAACCTGGATAAACACAGAGGAAA 2200
 QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
 Db 2201 GGCAGCAGACCCCTGCACCTACTGCTGCTGCCCTCCAGACAAATGCGAGTGCCTCAAGTTGCTC 2260
 QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
 Db 2261 CTGGGGGAAGGCTTCATCGAGATGCAACAGAGTCCAGAGAGACTCCCTGACACATT 2320
 QY 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuLeuThrGlnAlaLeuSerGlyArg 680
 Db 2321 GCCAAGCGCTCAGCAGCAGCAGCTGTGAGAGCTGTGACCCAGCCTTATCTGGGAGA 2380
 QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
 Db 2381 TTTAATTCTCAGTTTCACTGTTGAATATGAATGGCGACTACTCCACGAAGACTGGATGAA 2440
 QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
 Db 2441 AGTGATGACGACATGGATGAGAAATGTCAGCGCCAGTCCCAACCGCGGGAAGACCGGCC 2500
 QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
 Db 2501 ATCAGCTTCTACAGCTGGGTCTCAACAGCTTACGTACAGCTGTATCTTTGGCCAGA 2560
 QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
 Db 2561 GATGCTGCAAACTTGCCAGGAGAGCAGAGGGGCTTTTCATGCCAGCATCTTCAGAAAT 2620
 QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
 Db 2621 GAGACTTACGAGAGCCCTCTCAGTGGCAGCCACCTCCCGCCAGCTGCAAGCCCGCCAGC 2680
 QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
 Db 2681 ACCACGAGCCCGCCCGCTTCTCCACGGAATGTGGCAAGTTCAGACAGCTCTCTCT 2740
 QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
 Db 2741 GCTAACCCCTGTGGAAGACAACTCTGTAAGTGGAGCGTGGAGCGGAGCGGAGATCT 2800
 QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
 Db 2801 TCGTCAGATCCGCGAGCTGCTCATCCACCGCTGCCCTCTTCTCGCGTGACATCTACCAAT 2860
 QY 841 ProLeuThrProThrProProValAlaLysThrProSerValMetGluAlaLeu 860
 Db 2861 CCCCTGACCCCGCAGCCCGCCCGCTTCCCAAGAGCGCCAGCGTAATGGAAGCCTTG 2920
 QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880

Db 2921 AGCCAGCGGAGCAAGCCTGCCCGCCTGGGATCTCACAGATCAGGCCCCACCTCTGCC 2980
 QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
 Db 2981 CCAAGCGCGCCAGCGCCTCCCGCAGNAGAGCCTGCGCCGGGGCTGACAGATCCACC 3040
 QY 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
 Db 3041 CCACTGACCAACAAGGCCAACCGAGAGACCTGTGGATCTCTCTGCAACGGAAGCTCTG 3100
 QY 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
 Db 3101 GGTCTCTGTCCAATGTATGCTCTGCAGCCCTGCACCATGCTCTAGGAAGTCGAG 3160
 QY 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysAlaLysAspAsnPro 960
 Db 3161 GCAACCAAGTTGAAGCCTTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAACCC 3220
 QY 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
 Db 3221 GATGAGCTCACTTCTCCGAGGGGATGTGATCATCTGTGGACGGGAGGAGGACGAGAG 3280
 QY 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
 Db 3281 TGGTGGATTGGCCACATTTGATGGAGATCTCTGTGCGAAAGCGCATTCCTCCGCTGTCTTT 3340
 QY 1001 ValHisPheIleAlaAsp 1006
 Db 3341 GTGCACCTTTATCGCTGAC 3358

RESULT 2
 BC063308
 LOCUS
 DEFINITION Homo sapiens development and differentiation enhancing factor 2,
 mRNA (cDNA clone MGC:71731 IMAGE:30345039), complete cds.
 ACCESSION BC063308
 VERSION BC063308.1 GI:38648786
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 5544)
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Hopkins,R.F., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Uudin,T.B., Toohiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McSwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalek,U., Smaluk,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 5544)
 Strausberg,R.
 Direct Submission
 Submitted (02-DEC-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

TITLE
 JOURNAL
 PUBMED
 REFERENCE
 2 (bases 1 to 5544)
 Strausberg,R.
 Direct Submission
 Submitted (02-DEC-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Dr. Stefan Hansson
 cDNA Library Preparation: Michael Brownstein / Ted Usdin
 Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E.B. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 133 Row: h Column: 11
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4502248.
 Location/Qualifiers
 1. .5544
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:71731 IMAGE:30345039"
 /issue_type="Placenta, normal"
 /clone_lib="NIH MGC_147"
 /lab_host="DH10E"
 /note="Vector: pBluescript"
 1. .5544
 /gene="DDEF2"
 /note="synonyms: KIAA0400, PAP, SHAG1, PAG3, Pap-alpha"
 /db_xref="LocusID:8853"
 /db_xref="MIM:603817"
 294..3179
 /gene="DDEF2"
 /codon_start=1
 /product="DDEF2 protein"
 /protein_id="AAH63308.1"
 /db_xref="GI:38648787"
 /db_xref="LocusID:8853"
 /db_xref="MIM:603817"
 /translation="MPDQISVSSEFVAETHEDYKAPTASSFTTRTAQCRNTVAALBEAL
 DVDMVLYMKKSVKAINSGLAHVNEBQYQALEKFGGNRCVDRDPLGSAFTKFS
 VFTELTAFLKNLIQNMNNIISPLDLSLLKGLGVKGLKPPDKAMKDYKVTIKTLQ
 EKKEKHAHLGHIQTEISGAETAEEMEXERRFQLOCEVLLKYNELKIKYVDVLQ
 NLKYPHAQCNPFDGKLKAVESLKPSIETLSLDTITKQADEERQLILRDILKSA
 LQVEDSQIRQSTAYSLHQPKQNEHGTENGSLYKSGDGIKRWQKRCVKNGF
 LTSHGTANRPAPKALLTCQVKNPEEKCFDLISHDRTHYFAEDSQEQIWNMSVL
 QNSKEALNNAFNGDITGNNIIVQELTKIIESEVMTGNDVCCDCAPOPTWLSTN
 LGILTECSGIIHREIIVHYSRMQSLTDLVLGTSLLAKNIKAGNEIMECCLPAP
 DSVKPNPDSMNARKDYITAKYIERYARKKADNNAKLHSLCEAVKTRDIFGLQAY
 ADGVLDTEKPLANGHEPDTALHLAVRSVDRTSLHIVDLFVQNSGNDKQTKGSTA
 LHYCDLTDNAECLKLLRKGASTIEIANSSEGETPLDIARLKHCEELITQALSGFN
 SHYVVEYEWRLDEHDESDMDKXLPSPNRRDEPTISFYQLGNSLQSNVSLAR
 DANAANKDKORAPMBISILQNETYALLSGSPAPAPASPSTAPPLPRNVGDKPLT
 PTPPPVAKTPSWMEALSPKSPAPPGISQIRPPLPPQPPSRLPKKPAAGAKSTP
 LTNKQPRGVDSIALTEALGPLSNMLQPPAPMPEKRSQATKLPKPKVALYNVCADN
 PDELTFSEGVIIIVDGEEDQEWIIGHIDGDPGRKGAPFVSFVHFIAID"
 1209..1490
 /gene="DDEF2"
 /note="PH; Region: Pleckstrin homology domain"
 /db_xref="CDD:smart00233"
 1554..1898
 /gene="DDEF2"
 /note="ArfGAP; Region: Putative GTPase activating protein
 for Arf. Putative zinc fingers with GTPase activating
 proteins (GAPs) towards the small GTPase, Arf. The GAP of
 ARD1 stimulates GTPase hydrolysis for ARD1 but not ARFs"
 /db_xref="CDD:pfam01412"
 2052..2312
 /gene="DDEF2"
 /note="ANK; Region: ankyrin repeats"

REMARK
 COMMENT

FEATURES
 source
 gene
 CDS
 misc_feature
 misc_feature
 misc_feature

misc_feature
 2991..3170
 /genes="DDEF2"
 /notes="SH3; Region: Src homology 3 domains"
 /db_xref="CDD:smart00326"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.19e-176 Length: 5544
 Score: 4968.50 Matches: 955
 Percent Similarity: 95.13% Conservative: 2
 Best Local Similarity: 94.93% Mismatches: 4
 Query Match: 94.23% Indels: 45
 DB: 9 Gaps: 1
 US-09-914-042-1 (1-1006) x BC063308 (1-5544)
 Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
 Db 294 ATCCCGGACAGATCTCCGTGTCGGAATTCGTGGCGGAGACCCATGAGACTACAAAGGG 353
 Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 Db 354 CCACGGCCTCCAGCTTCACCCGCCACGCGCGAGTGC CGGAACACTGTGTGGCGCCATC 413
 Qy 41 GluGluAlaLeuAspValAepArgMetValLeuTyrLysMetLysLysSerValLysAla 60
 Db 414 GAGGAGGCTTTGGACGTGGACCGGATGGTCTCTTTACAAATGAAGAAATCCGTGAAAGCA 473
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrcInAlaLeuGlu 80
 Db 474 ATCAACAGCTCTCGGTGGCTCACGTGCGGAAAATGAAGACGACGTACACCCAGGCTCTGGAG 533
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
 Db 534 AGTTTGGCGGCAACCGTGTATGCAGAGATGACCCAGATTTAGGAAGTTCGTCTCTGAAG 593
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
 Db 594 TTCTCAGTGTTCACAAAGGAGTTGACAGCACATTTTCAAAAACCTGATTGAGAAATATGAAC 653
 Qy 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
 Db 654 AACATAATCTCTTCCTTTGGACAGTTCGTGTAAGGGGAGCTGTAAGAGGAGTGAAAGGG 713
 Qy 141 AspLeuLysPheProPheAspLysAlaTlpLysAspTyrGluThrLysIleThrLysIle 160
 Db 714 GATCTGAAAANGCCTTTTGATAAGCTTGAAGGACTATGAAACAAAAATATACCAAGATA 773
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
 Db 774 GAAAAGGAGAAAAGGAACACGCAAGCTCCATGGGATGATTCGGACTGAAATATAGCGGA 833
 Qy 181 AlaGluIleAlaGluLysMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
 Db 834 CGCGAAATTCGCAAGAGATGGAAGAGAGAGAGGCGCTTCTTCAGACTACAGATGTCCGAG 893
 Qy 201 TyrLeuLysValAsnGluLysIleLysLysLysLysValAspLeuLeuGlnAsnLeu 220
 Db 894 TATCTGCTGAAGGTCAACGAAATCAAGATTAAAAGGGAGTAGATTACTTTCAGAACTCTG 953
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
 Db 954 ATCAAAATACCTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAGCGCTGGAAGAG 101
 Qy 241 LeuLysProSerLysGluThrLeuSerThrAspLeuHisThrLysGlnAlaGlnAsp 260
 Db 1014 CTCAAACTTCCTTGAACCGCTGTCTACGGATCTTTCACACGATCAACAGGCGCCAGGAT 107
 Qy 261 GluGluArgGlnLeuLysGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
 Db 1074 GAAGAAAGAGGCGAGTTGTATACAGCTTCGAGATATTTTTGAATCCCGATTTCGAGGTGAA 113

QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1134 CAGAAAGAGGACTCCCAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGGA 1193
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1194 AACAAAGAACATGGGACCGAGCGGAACCGCAGCCTCTACAGNAGAGTGCAGGATCCGA 1253
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1254 AAAGTGTGGCAGAAAGGAATGTTTCAGTAAATAATGGTTTTCTGACCATATCCCATGGT 1313
QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1314 ACCGCTAACCGGCTCTCGCAAGCTCAACCTGTCAACTGCGAGTGAAGACCAACCTCT 1373
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1374 GAGNAGAAGAGTGTCTTGACCTTATTTTCACATGACAGAACTTACCACTTTCCAGCTGAA 1433
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1434 GATGAACAGGAATGTCAATAATGATGTCTGTGCTGCAAAATAGCAAAAGAAAGCTTTA 1493
QY 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1494 AACATGCAATTTAAGGGGATGACAAATATCTGGAGAAATAACATCGTCCAAGAACTGACA 1553
QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1554 AAGGAGATCATCTCAGAAGTGAGAGGATGACGGCAATGACGTCTGCTGTGACTGTGGG 1613
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1614 GCGCAGATCCTACATGCTTTCCACCAACCTGGGCATCCTGACCTGAGTGTCC 1673
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1674 GGAATCCACCGAGAGCTGGGGTTCATTTATTCAGGATGCAGTCCCTGACCTTAGATGTA 1733
QY 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1734 CTGGGAACATCTGAGCTGCTGCCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC 1793
QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1794 ATGGAATGTTGCCCTACAGCTGAGACTCAGTCAACCCACCCAGCGACATGAT 1853
QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHis 540
Db 1854 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGAGATACGCAAGGAAGACAC 1913
QY 541 AlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1914 GCGGATAACGCGGGAAGCTTCACAGTCTTTGCGAGGCGCTCAAAACGAGAGATATTTT 1973
QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1974 GGATGTCTCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCAAAC 2033
QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2034 GGACATGAGCGGATGAACCGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2093
QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2094 CTTCAATGTGTAGCTTTTATGTTCAGAACACAGTGGGAACCTGGATTAACACAGACGGAAA 2153
QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2154 GGCACACAGCCCTGCATCTACTGCTGCTGACCGCAATGCGAGTGCCTCAAGTTGCTC 2213
QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660

Db 2214 CTGCGGGGAAGGCGCTCCATCGAGATAGCAACAGAGTCAGGAGACATCCGCTGGACATT 2273
QY 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2274 GCCAAGCGCCTCAAGCACAGACACTGTGTGAGAGCTGTGACCCAGCCTTATCTCGAAGA 2333
QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db 2334 TTTAATTCTCAGCTTACGTTGAATATGAATGGCGACTACTCCACGAAGACCTCGATGAA 2393
QY 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2394 AGTGATGACGATGATGAGAAATTTGAGCGCCAGTCCCAACCCGCGGGAAGACCGGCC 2453
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2454 ATCAGCTTCTACAGCTGGGCTCCNACCAGCTTCACTCTACGCTGTATCTTTGGCCAGA 2513
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2514 GATGCTGCNAACCTTGCCAGAGCAAGCAGAGGGCTTTTCATGCCCAGCATCTTCAGAAAT 2573
QY 761 GluThrTyrGlyAlaLeuLeuSerClySerProProProAlaGlnProAlaIleProSer 780
Db 2574 GAGACTTACGAGCGCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGCAGCCCCCAGC 2633
QY 781 ThrThrSerAlaProLeuProProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2634 ACCACCGCGCCCCCGCTTCTCCACGGAAATGTTGGCAAA----- 2675
QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2675 ----- 2675
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2676 -----GAT 2678
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2679 CCCCTGACCCCAAGCGCCCGCCCGCTTGCACAGAGCCCGAGCTTAATGGAAAGCTTG 2738
QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2739 AGCCAGCGAGCAAGCCTGCCCGCTGGATCTCACAGATCAGCGCCCGCCACCTCTGCC 2798
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2799 CCACAGCGCCCGCAGCGCTCCCGCAGAAGAAGCCTGCGCGCGGGGCTGACAAGTCCACC 2858
QY 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 2859 CCACGTACCAACAAGGCCAACCCGAGAGACTGTGGATCTCTGTCAACGGAAGCTCTG 2918
QY 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 2919 GGTCTCTGTCCAATGCTATGCTCTGCTGAGCCCGCTGCACCCATCCCTAGGAAGTCGAG 2978
QY 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 2979 GCAACCAAGTTGAAGCCTTAAGCGGTGAAGCGCTCTATAACTGTGTGTGGCTGACAACCCC 3038
QY 961 AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu 980
Db 3039 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGGAGCGGGAGAGACCGAGAG 3098
QY 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3099 TGGTGGATTGGCCACATTTGATGGAGATCTCTGTGCGNAAGGCGCATTCGCCGTGTCTTT 3158
QY 1001 ValHisPheIleAlaAsp 1006
|||||

```

Db      3159 GTGCACCTTTATCGTGAC 3176
RESULT 3
AKI172944
LOCUS      AKI172944      4713 bp      mRNA      linear      ROD 28-JUL-2004
DEFINITION Mus musculus mRNA for mKIAA0400 protein.
ACCESSION AKI172944
VERSION    AKI172944.1 GI:50510472
KEYWORDS   FLI_CDNA.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Koseki,H.,
Hiraoka,S., Saga,Y., Seino,S., Nishimura,M., Kaisho,T., Hoshino,K.,
Kitamura,H., Nagase,T., Ohara,O. and Koga,H.
Prediction of the Coding Sequences of Mouse Homologues of KIAA
Gene: IV. The Complete Nucleotide Sequences of 500 Mouse
KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences
of cDNA Clones Randomly Sampled from Size-Fractionated Libraries
DNA Res. 11, 205-218 (2004)
2 (bases 1 to 4713)
Okazaki,N., Kikuno,R.F., Nagase,T., Ohara,O. and Koga,H.
Direct Submission
Submitted (19-MAY-2004) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics; 2-6-7
Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing: Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
LOCATION/Qualifiers
1. .4713
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mfj04247"
/note="vector:modified pBC SK+"
1. .4713
/gene="mKIAA0400"
<65. .2977
/gene="mKIAA0400"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence=not experimental
/product="mKIAA0400 protein"
/protein_id="BAD3222.1"
/db_xref="GI:50510473"
/translation="GGGSGPRGVAAMPDQISVEFVAETHEDYKAPTASSFTTRTAQ
CRNTVAIEEALDVRMVLVYKMKSVKAINISGLAHVNEREQYTOALEKFGCNCVCRD
DPDLGSAFLSVFTKELTALFKLIQNMNIIIPPLDSLKGDLKGVKGLKKPFDK
AWKDYETIKIEKEKBEKHALGMRITETISGAETABEMEKERRFFOLQWCEYLLKVN
EIKVKGVDLLOQLKYPHACNPFQDGLKAVESLKPSITLSTDLTHTIQAQDEER
QLIQRLDKLSALQVEKESRDSQLRSTAYSLHQPCNGKEHGTERNGLYKSDGI
RKWQKRKCSWNGSLTISHGTANRPAPKLLTCQVNPENKRCFPLISHDRTYHF
QAEDQECQIWSVLQNSKEELANNAFKGDDNTGENTIVBELTEIISEVQRMTCNDV
CCDGAEPDTNMLNLGTLTCEGSIHRELGVHYSRMQSLTLDVLGTSLLAKNLIG
NAGFNEIMECCLPSEDPKPNPMDIARKDYITAKYMERVYARKKHADTAALKHSLC
EAVKTRDIFGLQAVADGVDLTKETPLANGHEPDETALHLAVRSVDRPSTLHIVDFLQV
NSGNDKQGTQGSSTALHYCCLTIDNAECLKLLRGKASIEIANESGETFLDIAKLKHE
HCELLTQALSGRNFVHVHVEYERLHLEDLDEDDDDVDEKLQPSPNRDRPVSFYQ
LGSSQFNQSVANLQDNTANLTKQKRGFGPSILQNETYCAILSGSPSSQIPSTTS
APPPLPRNVGKDLPTTTPPPVAKTSGTLEAMNPKSKSQPSTQSPKPPPLPPPPSR
LPQKPKASGTDKPTPLTNKGQPGPEASGPLSNMALQPPAPMPKRSQATSKPKRVK
ALYNCVADNPDELTFSEGDVIVDGEEDQEWIGHIDGEPKRGAFPVSFVHFHFIAD"
ORIGIN
Alignment Scores:
Pred. No.: 1.03e-167 Length: 4713
Score: 4727.00 Matches: 908

```

```

Percent Similarity: 91.77%      Conservative: 18
Best Local Similarity: 89.99%      Mismatches: 29
Query Match: 89.65%      Indels: 54
DB: 10      Gaps: 3
US-09-914-042-1 (1-1006) x AKI172944 (1-4713)
QY      1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db      101 ATGCCGGACCAAGATCTCCGTGTCGGAATTCGTGCCCGAGACCCATGAGGACTCAAGCGG 160
QY      21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAenThrValAlaAlaIle 40
Db      161 CCCAGGCGCTTACCTTACCACCGCGACGCGCGCAGTGCCTGCGGAACACCGTGGCGGCATC 220
QY      41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLysAla 60
Db      221 GAGGAGGCTTTGGACGTGGACCGGATGGTTCTCTACAAAATGAAGAAATCTGTGAAGGCA 280
QY      61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db      281 ATCAATATCTCTGGGCTGGGCACACGTGGAGAACGAGGAACAGTACACCCAGGCTCTGGAG 340
QY      81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db      341 AAGTTCCGTGGCACTCTGTGTGCAGAGATGATCCAGACTTAGGAGCGGCTTCTGAAG 400
QY      101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLleGlnAsnMetAsn 120
Db      401 TTTCTGTGTGTTCAACAAGGAGTTGACGGCACTCTTCAAAAACCTGATTCAGAACATGAAC 460
QY      121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db      461 AACATCATCTCTGTTCCCTTTGGACAGTTTGTCTGAAGAGATCTGAAAGGAGTAAAGGG 520
QY      141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db      521 GATCTGAAAAGGCCCTTTGATAAAGCTTGAAGGACTATGAAACAAAAATAACCAAAATA 580
QY      161 GluLysGlnLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLleSerGly 180
Db      581 GAAAGGAGAGAAAAGGAGCATGCCAAGCTCCACGGGATGATTCGTACTGAAATTAAGCGGG 640
QY      181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db      641 GCTGAGATTGACAGGAGAGATGGAGAAAGACGGTTCTTCCAGCTGCAGATGTGTGAG 700
QY      201 TyrLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db      701 TACCTGCTGAAGGTCAATGAAATCAAGTGAAAAAGGAGTGGATTGTCTTCAGAACTCTG 760
QY      221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db      761 ATCAAGTACTTTCACGCCCACTGCAATTTTTTCAGGATGGATTGAAAGCAGTAGTAAAGC 820
QY      241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db      821 CTCAAGCTTTCATCGAGACGCTCTCCACGACCTCCACACCATCAACACAGCCCGCAGAT 880
QY      261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db      881 GAGGAACGGCGACAGCTGATACAACTTCGAGATATTTTGAAGTTCAGCATTCGAGGTGAA 940
QY      281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297
Db      941 CAGAAAGAGTCTAGGAGAGACTCAAACTTCGCCAAAGTACAGGCTTACAGCTTACATCAA 1000
QY      298 ProGlnGlyAsnLysGluHisGlyThrGluArgAenGlySerLeuTyrLysLysSerAsp 317
Db      1001 CCTCAGGGAACAAGGAGCATGGACAGAGAGAGACCGGGAACCTCTACAGAGAGCGCAT 1060
QY      318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337

```


Db 1061 GGGATCCGGAAAGTGTGGCAGAGAGGAAGTGTTCGGTTAAAAATGGCTTCTCCACCATC 1120
 Qy 338 SerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
 Db 1121 TCCACGGCAGCTGCCAAACGGGCCACTGCCAAGCTCAACCTGCTAACCTGCCAGGTGAAG 1180
 Qy 358 ThrAsnProGluGluLysCysPheAspLeuLysSerHisAspArgThrTyrHisPhe 377
 Db 1181 ACCAATCCGAGGAGAGAAAGTGTTCGACCTCATATCAGTGCAGAGACGTACCACCTTC 1240
 Qy 378 GlnAlaGluAspGluGlnCysGlnLeuTyrMetSerValLeuGlnAsnSerLysGlu 397
 Db 1241 CAAAGCGGAGAGCAACAGAGATGTCTAGATATGGATGTCTGACTGCAGACAGCAAGGAA 1300
 Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnLeuValGln 417
 Db 1301 GAAGCTCTGAAACAACGCCCTTTAAGGGTGATGACAACACTGGAGAAAAATAACATGTCCAA 1360
 Qy 418 GluLeuThrLysGluLeuLysSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
 Db 1361 GAGCTGACCAAGGAGATCATCTCGAGGGTGCAGAGGATGACGGCAATGACGTGTGCTGC 1420
 Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyLeuLeuThrCysIle 457
 Db 1421 GACTGTGGGGACCAAGATCCGACGTGGCTCTCTACCAACCTGGGCATCTGACTTGCATC 1480
 Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477
 Db 1481 GAGTCTCTGGGATCCACCGGAGCTGGGGTTCATTACTCCAGGATGCAGTCCCTGACA 1540
 Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497
 Db 1541 TTAGATGTATTGGGAACCTCTGAGCTTCTGCTTGCCAAAGATATCGGGAATGACGGCTTT 1600
 Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517
 Db 1601 AATGAATATTGAGTGTTCCTACCGTCTGAGGACCCAGCTCAAGCCCAACCCAGGCAGT 1660
 Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537
 Db 1661 GACATGATTGCAAGGAGGACTACATCACGCCAAGTACATGGAGAGAGATACGCACGG 1720
 Qy 538 LysLysHisAlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557
 Db 1721 AAAAAGCATGCAGACACCGCGCGAAGCTCCACAGCCTTTCGAGCGCGTCAAGACCAGA 1780
 Qy 558 AspilePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577
 Db 1781 GACATTTTGGGTACTCCAGCTTATGCTGATGTGTGGACCTCACAGAGAAAAATCCCA 1840
 Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597
 Db 1841 CTGGCCNATGGCATGAGCCAGATGAGACAGCCCTCCATCTTGCAGTCAAGTCTGTGGAC 1900
 Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617
 Db 1901 CGGACTTCCCTTCACTTGTAGACTTCTCTGGTCCAGAACAGTGGGAACTCGATAAACAG 1960
 Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLysLeuThrAspAsnAlaGluCysLeu 637
 Db 1961 ACAGCAAGGGGAGCAGACAGCCCTGCACCTACTGCTGCCTGCAGACAGCAATGCTGAGTGCCTC 2020
 Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657
 Db 2021 AAGCTCTCTCTGGGGGAAAGCCCTCCATTGATAGCCATGAGTCAAGGAGAGACCCCG 2080
 Qy 658 LeuAspileAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuLeuThrGlnAlaLeu 677
 Db 2081 TTGGACATTTGCTAAGCGCTCAAGCACCAACACTGTGAGGAACCTCTCACTCAGGCCCTTG 2140
 Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTyrAlaArgLeuLeuHisGluAsp 697
 Db 2141 TCGGGGCGGTTTAACTCCACGTTTCACTCGAGTATGAATGGCGGCTCTGTCATGAGGAC 2200

Qy 698 LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGlu 717
 Db 2201 CTGATGAGAGCGACGACGATGTGGAGAGCTTACAGCTAGTCCCAACCGACGGGAG 2260
 Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737
 Db 2261 GACCGGCTGTAGCTTCTACAGCTGGGGTCCAGCCAGTTTCAGTCAATGCTGTGCT 2320
 Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIle 757
 Db 2321 TTGGCCAGAGATCTGCAAACTCACCAGGACAGCAGAGGGGCTTTGGGCCAGCATC 2380
 Qy 758 LeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAla 777
 Db 2381 TTGCAGATGAGACCTATGGAGCCATCTGAGTGGCAGCCGCCCTCTCCAGTCAATA 2440
 Qy 778 AlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThr 797
 Db 2441 CCCCCAGCACCAACGAGCCGCCCTCCACCTCGGAATGTTGGCAAA 2491
 Qy 798 AlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArg 817
 Db 2491 2491
 Qy 818 GlnArgSerSerSerAspProProAlaValHisProProLeuProProLeuArgValThr 837
 Db 2491 2491
 Qy 838 SerThrAsnProLeuThrProThrProProProValAlaLysThrProSerValMet 857
 Db 2492 -----GATCCCTGACCAACACACACACCCCTCTCTGTGGCTTAAGACACTCTA 2545
 Qy 858 GluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProPro 877
 Db 2546 GAAGCATGAACACGCGAGCAGTCTCTCCAACTGGGACCTCACAGAGCAAGCCCCG 2605
 Qy 878 ProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAsp 897
 Db 2606 CCCCTGCCACCTCAGCCACCCAGCGCTGCTCAGAAAGAAACACAGTTCGGGAGTGC 2665
 Qy 898 LysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThr 917
 Db 2666 AAGCCAAACCCGCTGACCAACAAAGGACAGCGAGAGGACCT----- 2707
 Qy 918 GluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArg 937
 Db 2708 GAAGCTTCGGGTCTCTGTCCAACGCCATGGCCCTGCGAGCCCCCAGCACCATGCTAGG 2767
 Qy 938 LysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAla 957
 Db 2768 AAGTCACAGGCGACCAAGTCAAGCCCAAGCGAGTGAAGACACTCTACAACTGCGTGGCT 2827
 Qy 958 AspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGlu 977
 Db 2828 GACAACCCAGACGAGCTGACTTCTCTGAGGGCGACGTCATCATTTGTGTGGTGGGAAGAG 2887
 Qy 978 AspGlnGluTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhePro 997
 Db 2888 GATCAGGAGTGTGTGATCGGCCCATCATGATGAGAGCCCGCCGCAAGGAGCGTTTCCC 2947
 Qy 998 ValSerPheValHisPheIleAlaAsp 1006
 Db 2948 GTATCGTTGTGCACTTTATTGCTGAC 2974

RESULT 4

BC070750

LOCUS

DEFINITION

Xenopus laevis hypothetical protein MGC83760, mRNA

MGC:83760 IMAGE:6640628), complete cds.

ACCESSION

BC070750

VERSION

BC070750.1

KEYWORDS

MGC.

3989 bp mRNA linear VRT 19-JUL-2004
 Xenopus laevis hypothetical protein MGC83760, mRNA (cdna clone)

SOURCE	Xenopus laevis (African clawed frog)	
ORGANISM	Xenopus laevis	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Xenopus.	
REFERENCE	1 (bases 1 to 3989)	
AUTHORS	Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W. and Richardson, P.	
TITLE	Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative	
JOURNAL	Dev. Dyn. 225 (4), 384-391 (2002)	
PUBMED	12454917	
REFERENCE	2 (bases 1 to 3989)	
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Frange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Muliyil, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
PUBMED	12477932	
REFERENCE	3 (bases 1 to 3989)	
AUTHORS	Klein, S. and Strausberg, R.	
TITLE	Direct Submission	
JOURNAL	Submitted (10-MAY-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA	
REMARK	NIH-MGC Project	
COMMENT	Contact: XGC help desk Email: c9apbs@mail.nih.gov Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu, Parvaneh Saeedi, JR Santos, Angelique Scherch, Ursula Skalska, Duane Smalhus, Jeff Stott, Miranda Teal, George Yang, Jacques Schein, Asim Siddiqui, Rob Holt, Marco Marra.	
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAX Plate: 158 Row: f Column: 6 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein. Location/Qualifiers 1..3989 /organism="Xenopus laevis" /mol_type="mRNA" /db_xref="taxon:8355" /clone="MGC:83760 IMAGE:6640628" /tissue_type="Oocytes"	
source		
gene	/clone_lib="NICHD_XGC_001" /lab_host="DH10B" /note="Vector: pCMV-SPORT6.1" 1..3989 /gene="MGC83760" /db_xref="LocusID:431962" 369..3458 /gene="MGC83760" /codon_start=1 /product="MGC83760 protein" /protein_id="AAH70750.1" /db_xref="GI:47123082" /db_xref="LocusID:431962" /translation="MPDLISVSEFVCEVNEQYKPTASFTNRLSQCRNTVVAIDEAL EVDHTILNKKKSKVAIHTSGLTHVENEQYTALEKFGNSCVCDPDDLTGATFKY EVEHTILNKKKSKVAIHTSGLTHVENEQYTALEKFGNSCVCDPDDLTGATFKY EKEKHAHAKHGMFTIRISGABIAEEMEKERRVFOQMCEYLLKNEIKIKGVDLLQ NLIKYFHAQCNFODGLKAVETLAPSKSEKSLDHLTIKQVDEERKQLDSQRLDKAA LOVKEDSQRQSTAYSLHPOQNGKHGTEDGYLYKSGDGRKWKQKRCQTKVNGF LTISHGTANTPPAKNLTLTQVKTNPBEKKSFDLISHDRTYHFOADEAEACQIWMVSL QNSKEBALNNAFKGDSHAGENNIIVQELTKELISIQRMFGNDVCDGASPTWLSTN LGILITCECSIHRELGHVYSRIQSLTLDVLGTSELKLNKNGNSGFNEVMEACLPAD DSLKPTPCSDNNARKDYITTKYIERKARYKHNDKSLKGLCDVAKSRDIFALVHLH AEGVDLTAEPPLANGHEQGETALHVAVSVDNRSLHIVDFLVAQNSGLDKQTGSTA LHYCLTDNAECLLLRKGASTDIANDAGETPLDIAKRLKHAQCEDLIIQAMGRFN SHVHVEVRLHVLDESDDDDDEKMQQSPIRRERPVSYFOLGINQGPSVACLA REMAIAKDRPNFLPNIQNETYGAIVGNSNPGNAAGPATTPAPLPRTNTIKVHS TIFGSSIHHTPSNTPAWKASSIAESRQSSSPDSFIQPPVPLRVTSTATPPFP TKGPPALVTASLPAADPFSETTSQSGPKSKLTPSQPPAPMRKSQLPKMPKRV KAIYHCSADNPDELTFSEGEIIVVDGEDEKEMWICHIDGEDNRRGAFVPSVPHFIAE"	Alignment Scores: Pred. No.: 1.04e-148 Length: 3989 Score: 4215.50 Matches: 806 Percent Similarity: 86.42% Conservative: 85 Best Local Similarity: 78.18% Mismatches: 113 Query Match: 79.95% Indels: 27 Gaps: 5
ORIGIN	US-09-914-042-1 (1-1006) x BC070750 (1-3989)	
Qy	1 MetProbapGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20 	
Db	369 ATGCCGATCTGATCAGTGTCGGAATTTGTGCGAGACCAATGAGGATTACAGTCC 428 	
Qy	21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 40 	
Db	429 CCCACTGCTCCAATTTACCAACACAGGCTGAGCCAGTGCAGGAACACGGTCTGTGCTATA 488 	
Qy	41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60 :::	
Db	489 GATGAGGCTCTGGAGTTGATCACACGATCTCTGACACAAATGCAAAAATCTGTGAAAGCC 548 :::	
Qy	61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80 :::	
Db	549 ATTACACTTCTGGTCTGACCATGTGGAATATGAGGACAGTACACTCAAGCACTGGAA 608 :::	
Qy	81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100 :::	
Db	609 AAATTCGGAACCACTGCTTTGTAGGACACCTGATTTGGGAACGGCTTCCTCAAA 668 :::	
Qy	101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMetAsn 120 :::	
Db	669 TTCTCTGTCTTACCACAGGAGCTAATCTGACCTTTTCAAAAACCTGTGTGTCAGAACATGAAT 728 :::	
Qy	121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140 :::	
Db	729 AACATCATCATCTTCCACTTTCACAGCTTACTTGAAGGAGATCTCAAGGGGGTGAAGGG 788 :::	
Qy	141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160 :::	

Query Match:	73.53%	Indels:	197
DB:	10	Gaps:	3
US-09-914-042-1 (1-1006) x BC080847 (1-5072)			
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla	20
Db	289	ATGCCGACACAGATCTCCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTTACAAAGCGC	348
Qy	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	349	CCCAGGCTCTAGCTTCACACGCGCAGCCAGTCGCGGAACACCGTGGCGGCATC	408
Qy	41	GluGluAlaLeuAspValAspArgMetValIleuTyrIysMetLysLysSerValIysAla	60
Db	409	GAGGAGGCTTGGACGTGGACCGGATGTTCTCTACAAAATCAAGAAATCTGTGAAGCA	468
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu	80
Db	469	ATCAATATCTCTGGGCTGGCACACGTGGAGAACGAGAACAGTACCCAGGCTCTGGAG	528
Qy	81	LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys	100
Db	529	AAGTTCGGTGGCAACTGTGTGCAGAGATGATCCAGACTTAGGAAGCGCGTTTCTGAAG	588
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	589	TTCTCTGTGTTCACCAAGAGTTGACGCACCTCTTCAAAAACCTGATTCAGAACATGAAC	648
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly	140
Db	649	AACATCATCTCGTTCCTTGGACAGTTTGTGAAGAGAGATCTGAAAGGAGTAAAGGG	708
Qy	141	AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrIysIle	160
Db	709	GATCTGAAAAGCCCTTTGATTAAGCTTGGAGGACTATGAACAAAATAATACCAAAATA	768
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
Db	769	GAAGAAGGAGAAAAGGAGCATGCCAAGCTCCACGGGATGATTCTGTACTGAAATAGCGGG	828
Qy	181	AlaGluIleAlaGluLysGluMetGluLysGluArgPheGlnLeuGlnMetCysGlu	200
Db	829	GCTGAGATTGAGAGGAGATGGAGAAAGAAAGACGGTTCTTCCAGCTGCAGATGTGTGAG	888
Qy	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	889	TACCTGCTGAAGTCAATGAATCAAGGTGAAGAAGGAGTGGATTGCTTCAGAAATCTG	948
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer	240
Db	949	ATCAAGTACTTTCACGCCCCAGTGCATTTTTCAGGATGGATTGAAAGCAGTAGAAAGC	1008
Qy	241	LeuLysProSerIleGluThrIleuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1009	CTCAAGCCTTCATCGAGACGCTCTCCACGGACCTCCACACCATCAAAACAGGCCCCAGGAT	1068
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	1069	GAGGAAACGGGACACCTGATCACTTCGAGATATTTTGAAGTGCAGATTGCGAGTGGAA	1128
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1129	CAGAAAGAGGACTCACACTTCGCCAAAGTACAGCTACAGCTTACATCAACCTCAGGGA	1188
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1189	AACAAGGAGCATGGAACAGAGAGGAACCGGAACCTCTCAAGAGAGAGCATGCGATCCGG	1248
Qy	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1249	AAAGTGTGGCAGAGAGGAGTGTTCCTGTTAAATAATGGCTTCCTCACCATCTCCACGGC	1308

Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1309	ACT-----	1311
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1311	-----	1311
Qy	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu	400
Db	1311	-----	1311
Qy	401	AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1311	-----	1311
Qy	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1311	-----	1311
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1311	-----	1311
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
Db	1311	-----	1311
Qy	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1312	-----CTTGCCAAAGAAATTCGGGAATTCAGGCTTTAATGAAATTT	1350
Qy	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
Db	1351	ATGGAGTGTTCCTACCGTCTGAGGACCCAGCTCAAGCCCAACCCAGGAGTACATGATT	1410
Qy	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis	540
Db	1411	GCAAGGAAGGACTACATCACAGCAAGTACATGGAGAGAGATATGCACGAAAAAGCAT	1470
Qy	541	AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe	560
Db	1471	GCCGACACCGCGGGAAGCTCCACAGCCTTTGGCAGGCGCTCAAGACCAGAGACATTTT	1530
Qy	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
Db	1531	GGTTCCTCCAAAGCTTATGCTGATGGTGGACCTGCACAGAGAAATCCCACTGGCCAAT	1590
Qy	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
Db	1591	GGGCATGAGCCAGATGAGACAGCCCTCCATCTTGCACTCAGATCTGTGGACCGGACTTCC	1650
Qy	601	LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys	620
Db	1651	CTTCACATGTGTAGACTTCTCGTCCAGAACAGTGGGAACCTGGATAAACAGACAGCGAAG	1710
Qy	621	GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640
Db	1711	GGCAGCAGCCCTGCACTACTGCTGCCTGACAGCAATGCTGAGTGCCTCAAGCTCCTC	1770
Qy	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660
Db	1771	CTCGGGGAAAAGCCTCCATTCAGATAGCAATAGTCCAGAGAGACCCCGTTGGACATT	1830
Qy	661	AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg	680
Db	1831	GCTAAGCGGCTCAAGCAGCAACACTGTGAGAACTGCTCACTCAGGCTTGTGCGGGCGG	1890
Qy	681	PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu	700
Db	1891	TTTAACTCCCATGTTACGTCGAGTATGAATGGCGGCTGCTGCACAGAGACTTGGATGAG	1950
Qy	701	SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro	720

```
|||||...|
1951 AGCCGACGATGTGGAGAGAGCTACAGCTAGTCCCAACCGACGGAGGACCGGCT 2010
Qy IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db :|||...|
2011 GTTAGCTTCTACAGCTGGGTCCAGCAGATTTAGTCCAATGCTGTCTTTGGCCAGA 2070
Qy AspAlaAlaAsnLeuAlaValGlyLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db GATACTGAAACCTCACCAGACAGCAGAGGGCTTTGGGCCAGCATCTTGAGAAT 2130
Qy GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db :|||...|
2131 GAGACCTATGAGCATCTCAGTGGCAGCGCCCTCTCCAGTCATACCCCCAGC 2190
Qy ThrThrSerAlaProProLeuProProArgGlnValGlyLysValGlnThrAlaSerSer 800
Db ACCACGACGACCCCGCTCCACCTCGGAATGTTGGCAAA----- 2232
Qy AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db ----- 2232
Qy SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db :|||...|
2233 -----GAT 2235
Qy ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db CCCCTGACCAACACACACCCCTCTCTGTGGTAAAGACATCAGGAACCTAGAAAGCCATG 2295
Qy SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db :|||...|
2296 AACACGCGGAGAGTCTCTCCAACTCTGGGACCTCACAGAGCAAGCCCCCGCCCTGCCA 2355
Qy ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db CCTCAGCCACCCAGCGCGCTGCCTCAGAAAGAACCCAGCTTCCGGAGCTGACAGCCACAC 2415
Qy ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db CCGCTGACCAACAAAGGACGACGCGAGAGGACCT-----GAAGCTTCG 2457
Qy GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db GGTCTCTGTCCAAGCCATGGCTCGAGCCCGCCAGCACCCTATGCTAGGAAGTCACAG 2517
Qy AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db GCGACCAAGTCAAGCCCAAGCGAGTGAAGCACTCTACAACTGGTGGCTGACCAACCCA 2577
Qy AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu 980
Db GACGAGCTGACTTTCTCTGAGGGCGACGTGATCATTTGGATGGGGAAGAGGATCAGGAG 2637
Qy TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db TGGTGGATCGCGCACATCATGATGGAGAGCCCAAGCCGCAAAAGGAGCGTTTCCCGTATCGTTT 2697
Qy ValHisPheIleAlaAsp 1006
Db GTGCACCTTATTTGCTGAC 2715
```

```
RESULT 6
AR243296
LOCUS 2949 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 8 from patent US 6475778.
ACCESSION AR243296
VERSION AR243296.1 GI:27290454
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
```

```
REFERENCE 1 (bases 1 to 2949)
AUTHORS Roberts,T.M., King,F.J., Harris,D.F., Hu,E., Spiegelman,B. and Chan,J.
TITLE Differentiation enhancing factors and uses therefor
JOURNAL Patent: US 6475778-A B 05-NOV-2002;
FEATURES
source
1. .2949
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 4,19e-136 Length: 2949
Score: 3874.00 Matches: 755
Percent Similarity: 83.12% Conservative: 92
Best Local Similarity: 74.09% Mismatches: 122
Query Match: 73.47% Indels: 50
DB: Gaps: 13
US-09-914-042-1 (1-1006) x AR243296 (1-2949)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20
Db :|||...|
1 ATGCTCAGCCAGATAACAGTGGCGGAGTTGTCCAGGAGACAAATGAAGATTATAAATCG 60
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db :|||...|
61 CCCACCGCTCAAACTTCACCACCAGAAATGACTCACTGCAGGAACACAGTATCCGACTG 120
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db :|||...|
121 GAGGAGCCCTGGATGTGGACCGCAGTCTCTTTACAAGATGAAGAGTCAAGTCTAGGCT 180
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnThrThrGlnAlaLeuGlu 80
Db :|||...|
181 ATTTACGCTCGGCTCGGCTCATGTGAGAAATGAGGAGCAGTACACTCAAGCTCTGGAG 240
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db :|||...|
241 AAGTTCCGAGGAACTGTGTGTACAGAGATGACCCGACCTGGGATCAGCTTCTGAG 300
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db :|||...|
301 TTCTCCGCTTTCACCAAGGAGCTCAGCGCACTTTCAGAACCTGTTTCAGAACATGAAT 360
Qy 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db :|||...|
361 AATATCATTAACCTTCCCATTTGGACAGTCTGCTGAAGGAGATCTGAAAGGGGTTAAAGG 420
Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysIleThrLysIle 160
Db :|||...|
421 GATCTCAAGAGCCCTTCGATAAAGCTTGAAGAGCTACGAGACTTAAGTCTCTAAATA 480
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db :|||...|
481 GAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgPhePheGlnLeuGlnMetCysGlu 200
Db :|||...|
541 GCAGAGATAGCAGAGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db :|||...|
601 TACCTCTCAAGTCAATGAATCAAGATCAAAAGAGTGTGACCTGCTCCAGATCTC 660
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db :|||...|
661 ATCAATATCTTCCAGCAGCAGTGCACACTTCTTTTCCAGGATGCTCTCAAGAGCGGTG 720
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db :|||...|
721 CTCAAGCCCTCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
```

```
QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 781 GAAGAACGACAGACAGTAAACCACTTACGGATGTGCTAAAAAGCTCTGCAAGTGGAG 840
QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 841 CAGAAAGGAGGACTCTCAGGTTAGACAGAGCGCCACTACAGTCTGCACAGCGCGAGGCG 900
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 901 AACAAAGAGCATGGGACTGAGCGCAGCGCAACCTTTACAGAAAGAGTACGGGCTGCGG 960
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 961 AAAGTGTGGCAGAGAGAAAGTGCACAGTAAGAATGGATATTTGACCATCTCACATGGG 1020
QY 341 ThrAlaAsnArgProProAlaLysLeuHisLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1021 ACGGCAAAACAGACCTCCCGCCAAACTCAATCTTCTCACCTGTGAGTGAAGCACAACCCA 1080
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1081 GAGGAGAGAGAAAGTTTGTGACTCATCTCATGACAGAGAAATATCATTTCCAGGCAGAA 1140
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1141 GATGAGCCAGAGTGTCAAAATATGATCTCAGTGTCTGACAGACAGTAAAGAGAGGCGCTC 1200
QY 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnLeuThr 420
Db 1201 AACAAAGCGCTTCAAGGGCGCAGCATGTGTGTGAAATAAATGTTGTGCGAGGCTCACC 1260
QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1261 AAGGCCATCTGGGAGAGGTGAGCGGATGCGGGGAGAGATGTCTGTGCGAGTCCGGT 1320
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1321 GCTCCGGGCCCCACATGGCTCTCCACCAACCTGGGCATCTGACCTGCATCGAGTGTTCG 1380
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1381 GGGATCCACAGAGAGTGGGCGTCCATTACTCCCGAATCCAGTCCCTCACACTCGACGTC 1440
QY 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1441 CTCACACTCCGAGCTTCTGCTGCGCAAGAACGTTGGGGAATGTCTGCTTCAATGAGATC 1500
QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1501 ATGAGGCGCTGTCTGACGGCAGAGATGTGATCAAAACCGAATCCAGCCAGTGACATGCAG 1560
QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1561 GCGAGAGGAGGACTTATATGTCGCAAAATACACAGAGAAACGCTTCGCTCGTAAGAAGTGT 1620
QY 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1621 CCAGACGCGACTGTGAAAGCTGCACAGCTGTGTGATGCTGTGAAGCCCGGAGCATTTTC 1680
QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1681 TCTCTCATCCAGGCTATGCTGAAGAGTGGATCTGATGGAGCCCATTCCTCTGCGCTAAT 1740
QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 1741 GGACATGAACAAGGTGAGACGCGCTTTCATCTGGCCGTGAGACTGTGTGAGACAGAACTTCC 1800
QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 1801 CTACACATCATGAGCTTCTTCAACCAAAACAGTTTAAACCTGGATGAAGCAACGCGCTAAA 1860
QY 621 GlySerThrAlaLeuHisTyrCysLysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
```


Db	2053	CCAGACGCAGCTGTGCGAAGCTGCACACGCTCTGTGATCTGTGAAGCGCCGGGACATTTTC	2113
Qy	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
Db	2113	TCTCTCATCCAGGCTCTATGCTGAAGGAGTGGATCTGATGAGGCCCATCTCTCTGGCTAAT	2172
Qy	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
Db	2173	GGACATGAACAAGGTGAGACGGCTCTTCATCTGGCCGTGAGACTGTGTGGACAGAACTTCC	2232
Qy	601	LeuHisIleValAspPheLeuValGlnAsnSerClyAsnIleuAspIysGlnThrGlyLys	620
Db	2233	CTACACATCATCGACTTCCTCACCCAAAACAGTTTAAACCTGGATAGCAAAACGGGTAAA	2292
Qy	621	GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640
Db	2293	GGAGACACAGCTCTGCATTACTGCTGCTGACGGACAACGAGAGTGTCTCAAACTGCTG	2352
Qy	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660
Db	2353	CTCAGAGAAAAGCCTCATAGATATCGCTAAATGAAGCTGGAGAGACCCGTTGGACATC	2412
Qy	661	AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg	680
Db	2413	GCCAGGCGGACTCAACACATCTGCAGTGTGAGGAATCTGTAACACAGGCTCTTGACGGAAG	2472
Qy	681	PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu	700
Db	2473	TTCATGCTCATGTGCATGTGGAGTATGAGTGAGACTTCAGCATGAAGACCTGGACGAG	2532
Qy	701	SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro	720
Db	2533	AGTGATGAAGATCTGGATGAGAAGTCGAGTCCT-----CACCGCGGATGACGCGCC	2586
Qy	721	IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg	740
Db	2587	ATCAGCTGCTACACACCGGCAGTAACTCCCTTCAGCTGAGTCAGCCACGCTGAGCCGA	2646
Qy	741	AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleGlnAsn	760
Db	2647	GACGCTCGAGACCTGGTTAAAGACACAAACGC-----TTTGTGCCAAACCTGGTCAACAT	2703
Qy	761	GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer	780
Db	2704	GAACCTACGGACCATTAACACACAGCTCACCC-----GTCAGCCTG	2748
Qy	781	ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer	800
Db	2749	TCCTCTTCTGCTCCACCTCTACCAACCCGAAACCTA-----GTTACGCGCTCTGCTCTT	2802
Qy	801	AlaAsnThrLeu-----TrpLysThrAsnSerValSerValAspGlyGly	815
Db	2803	GCAGGACTGACTCAGGATCTCCCGGCTGGAAGCCTGGCTCTCTGGATCTGACGGCGC---	2859
Qy	816	SerArgGlnArgSerSerAspProAlaValHisProProLeuProProLeuArg	835
Db	2860	---AGACAGAGATCTCTCTGACCTCCCAACATGATCTCTCTGGCGCTCTCCCTTACGG	2916
Qy	836	ValThrSerThrAsnProLeuThrProThr---ProProProValAlaLysThrPro	854
Db	2917	GTCACTTCCACCTCCCTTCTAATGCCACGGGTGCTGCTCTCTCTGGCTAAAGCTACT	2976
Qy	855	SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProProGlyIleSer	872
Db	2977	GGTATGATGAGACCATGAATATGCAACCCAAACCCGACAGGGGGCTCTCTGACAGAAC	3036
Qy	873	GlnIleArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysPro	892
Db	3037	ATCAACCGG-----	3045
Qy	893	AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro	908

Db	3046	GCTACAGTGGCGGACAAAACTTCAGCAAAAGCACACTGTATGGCTCC-----	3093
Qy	909	ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet	927
Db	3094	--GGATCCATCGAGACACCAGCTAAAGAAGTCCCGAGGAGCCACAAAACACCACTGGT	3150
Qy	928	ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys	947
Db	3151	CAAACTCTGCCTCGCGACCCACATGCCCCAGAAA-----ACGTATTTCGAAGCCGAAG	3201
Qy	948	ArgValLysAlaLeuTyrrAsnCyeValAlaAspAsnProAspGluLeuThrPheSerGlu	967
Db	3202	CGTGTGAAGGCCATGATTAACCTGTGTGGCGAATAATCCAGACGAGCTGACTTCTCTGAG	3261
Qy	968	GlyAspValIleIleValAspGlyGluGluAspGlnGluTrrPrilleGlyHisIleAsp	987
Db	3262	GGAGACTTATCGTGTGGATGGAGAGGAGGACCAGGAGTGGTGGTGGCCACATTGAG	3321
Qy	988	GlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp	1006
Db	3322	GGAGAGCAATGAGAGAGGAGCGTTTCCTGTCTCACGTTGTACAGTTTCATTATGGAC	3378
RESULT 8			
LOCUS	I74317	4486 bp DNA linear PAT 03-APR-1998	
DEFINITION	Sequence 15 from patent US 5688641.		
ACCESSION	I74317		
VERSION	I74317.1 GI:3010458		
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 4486)		
AUTHORS	Sager, R.; Zou, Z.; Lee, S. Whan, and Tomaasetto, C. Laure.		
TITLE	Cancer diagnosis using nucleic acid hybridization		
JOURNAL	Patent: US 5688641-A 15 18-NOV-1997;		
FEATURES	Location/Qualifiers		
source	1..4486		
	/organism="unknown"		
	/mol_type="unassigned DNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	4.35e-129	Length:	4486
Score:	3691.00	Matches:	701
Percent Similarity:	99.29%	Conservative:	1
Best Local Similarity:	99.15%	Mismatches:	3
Query Match:	70.00%	Indels:	2
DB:	6	Gaps:	0
US-09-914-042-1 (1-1006) x I74317 (1-4486)			
Qy	302	LysGluHisGlyThrGluArgAsnGlySerLeuTyrrLysSerAspGlyIleArgLys	321
Db	1	AAGGAACATGGGACCGGACCGGAAGCGGAGCGCTCTACAAAGAAGAGTAGTCGGGATCCGAAA	60
Qy	322	ValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr	341
Db	61	GTGTGGCAGAAAAGGAATGTTCAGTTAAAAATGGTTTTCTTGACCATATCCCATGGTACC	120
Qy	342	AlaAsnArg-ProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGln	361
Db	121	GCTAACCGGGCTCTCTCAAAAGCTCAACCTGCTAACTGCCAGGTGAAGACCAACCTCGA	180
Qy	361	uGluLysCysPheAspLeuIleSerHisAspArgThrTyrrHisPheGlnAlaGluAs	381
Db	181	GGAGAAGAAGTGCTTTTGACCTTATTTTCATGATGACAGAACTTACCACCTTTCAAGCTGA	240
Qy	381	pGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeuAs	401
Db	241	TGAACGAGAATGTCMAATATGGATGTCTGTGCTGCAAAATAGCAAGAGAGCTTTTAA	300
Qy	401	nAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys	421


```
Db 301 CAATGCTTTAAGGGGATGCAATACTGGAGAAATAAATCGTCCAAAGAACTGACAAA 360
Qy 421 sGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAl 441
Db 361 GGAGATCATCTCAGAAGTGCAGAGATGACGGCAATACGCTGCTGTGACTGTGGGGC 420
Qy 441 aProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysValIleGluCysSerG1 461
Db 421 GCCAGATCCTACATGGCTTTTCACCACTCGGCATCTCTGACCTCGCATCGAGTGTTCGG 480
Qy 461 yIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLe 481
Db 481 AATCCACGAGAGCTGGGGTTCATTATTCCAGGATGCAGTCCCTGCATCTTAGATGACT 540
Qy 481 uGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnIleGlyAsnIleGlyAsnIleMe 501
Db 541 GGGAAACATCTAGCTGCTGCTCGCAAGATATTGGGAATGTCAGGCTTTAATGAGATCAT 600
Qy 501 tGluCysCysLeuProAlaGluAspSerValIysProAsnProGlySerAspMetAsnAl 521
Db 601 GGAATGTTGCCCTACAGCTGAGACTCAGTCAACCCCAACCCAGGACGACATGAATGC 660
Qy 521 aArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAl 541
Db 661 AAGAAGAGACTACATCAGCAAGCAAGTACATCGAGAGAGATACGCAAGAGAGACGACGC 720
Qy 541 aAspAsnAlaLysLeuHisSerLeuCysGluAlaValIysThrArgAspIlePheG1 561
Db 721 GGATAACGGCGCAAGCTTCAGTCTTTGCGAGGCGCTCAAAACGAGAGATATTTTGG 780
Qy 561 yLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluIysIleProLeuAlaAsnG1 581
Db 781 ATTGCTCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAATCCCACTGGCCAAACGG 840
Qy 581 yHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLe 601
Db 841 ACATGAGCGGAGTGAAGACGGCCCTCACCTTGCAGTCAGATCCGTGGATCGAACCTCTCT 900
Qy 601 uHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysG1 621
Db 901 TCACATTGTAGACTTTTGTAGTTTTCAGACAGTGGGAACCTGGATAAACAGACAGGGAAGG 960
Qy 621 ySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuLe 641
Db 961 CAGCACGCCCTGCACTACTGCTGCTGACCGACAATGCCAGTGCCTCAAGTTGCTCCT 1020
Qy 641 uArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAl 661
Db 1021 CGGGGGAGGCTCCATCGAGATAGCAATAGTTCAGAGAGACTCCGCTGGACATTGC 1080
Qy 661 aLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArgPh 681
Db 1081 CAAGCGCTCAAGCAGCAGCACTGTGAGAGCTGTGACCCCAAGCTTATCTGGAAGATT 1140
Qy 681 eAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGluSe 701
Db 1141 TAAATCTCACGTTTCACTGTAATATGAATGGGCACTTCTCCACGAGACCTTGGATGAAAG 1200
Qy 701 rAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgProI1 721
Db 1201 TGATACGACATGGATGAGAAATATGAGCCAGTCCCAACCGCGGGAGACCGGCCCAT 1260
Qy 721 eSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAs 741
Db 1261 CAGCTTCTACCACTGGGCTCCCAACGAGCTTCAGTCTAACGCTGATCTTTGGCCAGAGA 1320
Qy 741 pAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnG1 761
Db 1321 TGCTGCAAACTTGGCAAGGACAGCAGAGGCTTTTCATGCGCCAGCATCTTGCAGAAATGA 1380
Qy 761 uThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSerTh 781
```

```
Db 1381 GACTTACGAGCCCTCTCTGATGGCAGCCACCTCCCGCCAGCCTCGACGCCCCAGCAC 1440
Qy 781 rThrSerAla-ProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSerA 801
Db 1441 CACAGCGCCCGCCCGCTTCTCTCCAGGAATGTGGCAAGTTTCAGACACCTCTCTCTG 1500
Qy 801 laAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSers 821
Db 1501 CTAACACCTGTGGAAGACAACTCTGTAAAGTGTGGACGGTGGAGCGCGGAGCGATCTT 1560
Qy 821 erSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsnP 841
Db 1561 CGTCAGATCCGCCAGCTGTCCATCCACCGCTGCCCTCTTTCGGCTGCACATCTACCAATC 1620
Qy 841 roLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeuS 861
Db 1621 CCCTGAGCCCCACCGCCCGCCCTGTCGCAAGACCTGCCAGGGGCTGATGAAGACCTTGA 1680
Qy 861 erGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuProP 881
Db 1681 GCCAGCGCAGCAAGCTTGGCCGCTGGATCTTCAGAGATCAGGCCCGCCACCTCTGCC 1740
Qy 881 roGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrp 901
Db 1741 CACAGCGCCCGCAGCGCTCCCGCAGAAGAAGCTGCCGCCAGGGGCTGACAAGTCCACCC 1800
Qy 901 roLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuG 921
Db 1801 CACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCTGG 1860
Qy 921 lyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGlnA 941
Db 1861 GTCTCTGTGCATGTCTATGGTCTTCGAGCCCCCTGCACCCATGCTAGGAAGTCGCAGG 1920
Qy 941 laThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProA 961
Db 1921 CAACCAAGTTGAAGCTTAAGCGGTGAAAGCGCTCTATACTGTGTGCTGACAACCCCG 1980
Qy 961 spGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGluT 981
Db 1981 ATGAGCTCACCTTCTCCAGGGGATGTGATCATCGTGAGCGGGAGAGACCGAGGAGT 2040
Qy 981 rpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPheV 1001
Db 2041 GGTGGATGGCCACATGTATGGAGATCCTGTGTCGCAAGAGGCGCATTCGCGGTGTCAATTG 2100
Qy 1001 alHisPheIleAlaAsp 1006
Db 2101 TGCACCTTATCGCTGAC 2117

RESULT 9
I74314 4328 bp DNA linear PAT 03-APR-1998
LOCUS Sequence 12 from patent US 5688641.
DEFINITION I74314
ACCESSION I74314
VERSION I74314.1 GI:3010455
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 4328)
AUTHORS Seger,R.; Zou,Z.; Lee,S.Whan. and Tomasetto,C.Laure.
TITLE Cancer diagnosis using nucleic acid hybridization
JOURNAL Patent: US 5688641-A 12 18-NOV-1997;
FEATURES
source
1. 4328
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores: 6.98e-119 Length: 4328
Pred. No.: 3417.00 Matches: 648
Score:
```

Percent Similarity: 99.39% Conservatives: 1
 Best Local Similarity: 99.23% Mismatches: 3
 Query Match: 64.80% Indels: 1
 DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x I74314 (1-4328)

QY 355 GlnValLysThrAsnProGluLysLysCysPheAspLeuLeuSerHisAspArgThr 374
 DB 3 CAGGTGAAGACCAACCTTGAGGAGAGAGAGTCTTTGACCTTATTTACATGACAGAACT 62
 QY 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTropMetSerValLeuGlnAsn 394
 DB 63 TACCACTTTCAAGCTGAAGATGAACAGGAATGTCAATATGATGTCTGTGTGCAAAAT 122
 QY 395 SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsn 414
 DB 123 AGCAAGAAGAGACTTTAAACAATGCATTTAAGGGGGATGACAACTTGGAGAAATAAC 182
 QY 415 IleValGlnGluLeuThrLysGluIleLeuSerGluValGlnArgMetThrGlyAsnAsp 434
 DB 183 ATGCTCCAGAACTGACAAAGAGATCATCTCAGAAGTGCAGAGATGACGGGCAATGAC 242
 QY 435 ValCysCysAspCysGlyAlaProAspProThrTropLeuSerThrAsnLeuGlyIleLeu 454
 DB 243 GTCTCTGTGACTGTGGGGCGGCAGATCCTACATGGCTTCCACCAACTGGGCATCTCTG 302
 QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474
 DB 303 ACCTGCATCGAGTGTCCGGAAATCACCAGAGAGCTGGGGTTCATTTCCAGAGATGAC 362
 QY 475 SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsn 494
 DB 363 TCCCTGACCTTAGATGTAGTGGGAACATCTGAGCTGTCTCGCCCAAGAAATATGGGAAT 422
 QY 495 AlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsn 514
 DB 423 GCAGCTTTAATGAGATCATGGAATGTGTGCTTACCAGCTGAGGACTCAGTCAAAACCCAAC 482
 QY 515 ProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 534
 DB 483 CCAGCAGGACATGATGATGCAAGAAGAGACTACATCAGCCCAAGTACATCAGAGAGAGA 542
 QY 535 TyrAlaArgLysLysHisAlaAspAsnAlaLysLeuHisSerLeuCysGluAlaVal 554
 DB 543 TACGCAAGGAGAGACACCGGATTAACCGGGGAGCTTCAAGTCTTTTGGAGGCGCTC 602
 QY 555 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGlu 574
 DB 603 AAACGAGAGATATTTTGGATTTGCTCCAAGCTTATGCTGATGGTGTGGATCTTACGGAA 662
 QY 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594
 DB 663 AAAATCCCACTGCTCCACGACATGAGCGCGATGAAGCGCCCTCCACTTGCAGTCAGA 722
 QY 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeu 614
 DB 723 TCCGTGGATCGAACCTCTCTTCCACATTTAGACTTTTATGTTTCAAGACAGTGGGAACCTG 782
 QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634
 DB 783 GATAAACAGACAGAGGAAAGGACAGCAGCCCTGCACTACTGCTGCTGACCGACAATGCC 842
 QY 635 GluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654
 DB 843 GAGTGCCTCAAGTTGCTCTCGGGGGAGAGGCTCCATCGAGATAGCAAAATGAGTCAGA 902
 QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThr 674
 DB 903 GAGACTCCGCTGACATTTGCCAAGCGCTCAGACAGAGCACTGTGAGGAGCTGTGACC 962
 QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluThrPArgLeuLeu 694

DB 963 CAAGCCTTATCTGGGAAGATTTTAATTTCTCAGCTTCACTGTTGAATATGAATGGCAGCTACTC 1022
 QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714
 DB 1023 CACGAAGACCTGGATGAAAGTGTACGACATGATGAGAAATTCAGCCCACTGCCAAC 1082
 QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734
 DB 1083 CGCGGGAGAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCCAACAGCTTCACTCAAC 1142
 QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754
 DB 1143 GCTGTATCTTTGGCCAGAGATGCTGCAAACTTGGCAAGAGACAGAGGGCTTTTCATG 1202
 QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774
 DB 1203 CCCAGCATCTTGCAGAAATGAGACTTTACGAGGCCCTCTCTGAGTGGCAGCCACCTCCGCC 1262
 QY 775 GlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794
 DB 1263 CAGCCTGAGCGCCCGCAGCACCAGCCGCCCGCCCGCTTCTCCACGGAATGTTGGCAA 1322
 QY 794 sValGlnThrAlaSerSerAlaAsnThrLeuTropLysThrAsnSerValSerValAspG 814
 DB 1323 AGTTCAGACAGCTCTCTCTGTAAACCTGTGGAAGACAAACTCTGTAAGTGTGACGG 1382
 QY 814 yGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLeu 834
 DB 1383 TGGAAAGCCGCGCAGCGATCTTCTGTGATCCGCCAGCTGTCCATCCACCGCTGCCCTCT 1442
 QY 834 uArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysThrPr 854
 DB 1443 TCGCTGACATCTTACCAATCCCTGACCCCGCCCGCCCGCCCGCTTCCCAAGACGCC 1502
 QY 854 oSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGlnI 874
 DB 1503 CAGCGTAATGAAGCTTTGAGCCAGCCGAGCAAGCTCTGCCCGCTTGGGATCTCAGAT 1562
 QY 874 eArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPr 894
 DB 1563 CAGGCCCCACCTCTGCCCCCAGCAGCCCGCCCGCCCGCTTCCCGCAGAGAGAGCTTGGCC 1622
 QY 894 oGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeu 914
 DB 1623 AGGGGCTGACAAGTCCACCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGATCT 1682
 QY 914 uSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPr 934
 DB 1683 CTCTGCAACGGAAGCTCTGGGTCTCTCTGCAATGTATGGTCTTGCAGCCCGCTTGCACC 1742
 QY 934 oMetProArgLysSerGlnAlaThrLysLysProLysArgValLysAlaLeuTyrAs 954
 DB 1743 CATGCTAGGAGTTCGAGGCAACCAAGTTGAGGCTAAGCGGTGAAGCGCTCTATAA 1802
 QY 954 nCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAs 974
 DB 1803 CTGTGTGCTGACAAACCCCGATGAGTCACTTCTCCAGGGGATGTGATCATCGTGA 1862
 QY 974 pGlyGluGluAspGlnGluTropIleGlyHisIleAspGlyAspProGlyArgLysG 994
 DB 1863 CGGGAGAGAGACAGGAGTGTGATGGCCACATGATGAGATCTCTGGTGGCAAGG 1922
 QY 994 yAlaPheProValSerPheValHisPheIleAlaAsp 1006
 DB 1923 CGCATTCGCGGTGTCATTTGTGCACTTTATCGCTGAC 1959

RESULT 10

AR243298

LOCUS

Sequence 11 from patent US 6475778.

DEFINITION

AR243298

ACCESSION

VERSION

AR243298.1

GI: 27290456

KEYWORDS

AR243298 2712 bp DNA linear PAT 20-DBC-2002


```
Db 1798 CTCATATTGTGGACTTCTTGTACAGAACAGTGGCAATTTAGACAGACAGACCCAAA 1857
QY GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 1858 GGAAGCACAGCGCTACATTATTGCTTGTGCTGATGATGATGATGATGATGATGATGATG 1917
QY LeuArgGlyLysAlaSerIleGluLeuAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 1918 CTGCGGGGGAAGCATCTGTGACGATTTACTAATGATGCTGGAGAGACTGCTCTGATTTG 1977
QY AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 1978 GGGAGCGCTCTCAACACTCCAAATGCGAGGAGCTGCTGACTACGGCGCAGACGGGGAG 2037
QY PheAsnSerHisValHisValGluTyrGluThrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2038 TTCAATGTCATGTCATGTCGAATATGATGCTGGCTGTCATATGAGGATCTGGACGAG 2097
QY SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2098 AGCGAAGATGAGATGGAGACAAG-----CCATTCCCATCAGGCGTGAGGAGGTCCA 2151
QY IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2152 ATAACTGTATAGTTCAGGCGAGTGGCCCATGATGCCCAACATCAGGCGCTCTGGCTCG 2211
QY AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2212 GACGTGGCCAAATGTGTCAATAAAGCAGAGGGCTTTTATCCGAGCATGATGATGAAC 2271
QY GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer 780
Db 2272 GAGACTTACGGCACCATGCTCGATCCCACTCCACCACCTG-----GGTTTACCAGGA 2325
QY ThrThrSerAlaProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2326 GTACCTGGCATTCCTCTTTACCCCTCGCCCTTGGGAAGGGGA----- 2370
QY AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2371 -----TGG-----AGTCCACCAATGGAGACATCGGTAGACAGAGTCA 2409
QY SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2410 TGTTCAGATCCTGCA-----AACCCTCAAACTCCTGGAACAAAATAAATCTCTGTAT 2460
QY ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2461 GTTCTGCTCTGCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2493
QY SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2494 -----CCCAAGAGACTCCACTCCAGATCCAAAGGCCAGTCTTCTCTCCAGCACC 2547
QY ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2548 ACGGCTCTCT-----CCTGCACCA-----TCCGCA 2571
QY ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 2572 CGGCTCTTTATTCACCTGCTCTCTCAGGCCAGCGCCT----- 2610
QY GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 2611 -----GTAGTGCCCTCTGCACCAAGTTATGCCCACTTCGTC 2646
QY AlaThrLysLeuLysProLysArgValLysAla 951
Db 2647 CTGACTGATGTCAAAGTCTGCTGTCTAAAGCC 2679
RESULT 11
AR243297 4595 bp DNA linear PAT 20-DEC-2002
LOCUS
```

```
DEFINITION Sequence 9 from patent US 6475778.
ACCESSION AR243297
VERSION AR243297.1 GI:27290455
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4595)
AUTHORS Roberts,T.M., King,F.J., Harris,D.F., Hu,E., Spiegelman,B. and Chan,J.
TITLE Differentiation enhancing factors and uses therefor
JOURNAL Patent: US 6475778-A 9 05-NOV-2002;
FEATURES
    Location/Qualifiers
        source
            1..4595
            /organism="unknown"
            /mol_type="genomic DNA"
ORIGIN
Alignment Scores:
  Pred. No.: 4.41e-109 Length: 4595
  Score: 3155.00 Matches: 621
  Percent Similarity: 75.60% Conservative: 98
  Best Local Similarity: 65.30% Mismatches: 174
  Query Match: 59.83% Indels: 58
  DB: Gaps: 10
US-09-914-042-1 (1-1006) x AR243297 (1-4595)
QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 300 ATGCTGTATGACGTGACAGTGCAGTGTGTGGATATATTACCCATGAGGACTATAAGCA 359
QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 360 CCGCAACATCAGTCTTCTGCACGCGCATGCTCACTGCAGGAATACAGTCCGCCGCTCG 419
QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 420 GAAGAGGCGCTGGATCTGACCGCATGCTACTGCACAAAATGAAGAAGTCAGTCAAGGCC 479
QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 480 ATAAACAGCTCTGGTCAGACTCATGTAGAACAGAGGAGCAGTATACAGGCCCATAGAG 539
QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 540 AGGTTACGGATAACACTGTGTACAAAGATGACCTTGAGATGTCCTCAATTTCTCTCACA 599
QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 600 TTCGCTGTTTCCACCAAGGAGCTTACTGCTCTTTTCAAGAACTTCTGTACAGAACATGAAT 659
QY 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 660 AACATCATCACTTTTCCACTAGACAGTCTGCTAAAGGGAGACCTCAAGAGGATCAAGGG 719
QY 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 720 GATTGAAAAGAACCCATTGATTAAGCATGGAGGATTAATGAACCAACTGAGCAAGATT 779
QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 780 GAGAAAGAAAAGCGAGAACATGCCAAACAGCACGCTGTCTGATCCGAAACAGAGATCATGTA 839
QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 840 GGAGAGATCGCAGAGAGATGGAGAAAGAGAGACGCTGTTTTCAGCTTTTCAGATGTGTGAG 899
QY 201 TyrLeuLeuLysValAsnGluIleLysIleLysValLysGlyValAspLeuGlnAsnLeu 220
Db 900 TACCTCTTAAGTGAATGAATCAAGTCAAAAGGGGGGTCGACCTGCTTCCACACCTC 959
QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
```

Db |||||ATCAAACTTTTCATGCCAGTGCATTTCTTTCAGGATGGCTAAAGTCTGTGACAAAT 1019 |||||:|||||
Qy 241 LeuIysProSerIleGluThrLeuSerThrAspLeuHisThrIleIysGlnAlaGlnAsp 260 |||||:|||||
Db 1020 CTGAAACCTTTTCAGAAAGCTTTCACAGACTTAACCGCGAAACAAACAGACTCAAGAT 1079 |||||:|||||
Qy 261 GluGluArgGlnLeuIleGlnLeuArgAspIleLeuIysSerAlaLeuGlnValGlu 280 |||||:|||||
Db 1080 GCAGAAAGAAACAGTGTCTGCAGCTGAAAGAACTCTTAAATCTGTCTACAGTCTGAG 1139 |||||:|||||
Qy 281 GlnIysGluAspSerGlnIleArgGlnSerThrAlaIleSerLeuHisGlnProGlnGly 300 |||||:|||||
Db 1140 TGTAAAGAGGATGCTCAGTCAAAAGCAGAACCGCAGGCTACAGTCTTCACCAAGTTGCGAGGC 1199 |||||:|||||
Qy 301 AsnIysGluHisGlyThrGluArgAsnGlySerLeuIleIysSerAspGlyIleArg 320 |||||:|||||
Db 1200 AATAAGCTCAGCGCAGCGCTCTGGGATGCTCTCTCAACGCGCAGGAGGACTGAGG 1259 |||||:|||||
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340 |||||:|||||
Db 1260 AAAGTTTGCAGAAAGAAAGTGTCTGTGAAATGGAATGTTGTTGACTATTTCACATGGA 1319 |||||:|||||
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnPro 360 |||||:|||||
Db 1320 ACGCCCATGCACCGCCAGCAACCTGAACCTTTAACTGCCAAGTGAAGGTAAACCA 1379 |||||:|||||
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyRHisPheGlnAlaGlu 380 |||||:|||||
Db 1380 GATGAGAAAAATGCTTTGATCTCATATCATATGATGATGATGATGATGATGATGATGAT 1439 |||||:|||||
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400 |||||:|||||
Db 1440 GATGAGGAGAGTGTGAGTATGGGTCTGTCTCCAGAACAGTAAAGAAAGGCGCTG 1499 |||||:|||||
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420 |||||:|||||
Db 1500 AACATGCTTTAAAGAGCATCAGAAATGAGGAGAGAAATAACATGTTTCGAGAGTCACT 1559 |||||:|||||
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440 |||||:|||||
Db 1560 AAGGCATCTGTGGGGAAAGTGAAGAAATGAGCGCAATGAGCTGTGCTGACTGTGGA 1619 |||||:|||||
Qy 441 AlaProAspProThrLysSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460 |||||:|||||
Db 1620 GCTTCCAATCCAACATGGCTCTCCAAACACCTGGGTGTGTAATGTGCAATGAATGCTCT 1679 |||||:|||||
Qy 461 GlyIleHisArgGluLeuGlyValHisIleIleIleIleIleIleIleIleIleIleIleIle 480 |||||:|||||
Db 1680 GGGATCCATCGGAAATGGCGTCCACTACTCCCGAATACAGTCTCTGACACTGGACCTC 1739 |||||:|||||
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500 |||||:|||||
Db 1740 TTAGGCACATCTGAACATATTGCTGAACAGTGTGGGAAATGACAGCATTTCAATGAATC 1799 |||||:|||||
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520 |||||:|||||
Db 1800 ATGGAAGCAAACTGTTTCAGAG--ATCCCAAAACCCCTACCTTCTAGTGACATGAG 1856 |||||:|||||
Qy 521 AlaArgLysAspTyRileThrAlaLysTyRileGluArgArgTyRAlaArgLysLysHis 540 |||||:|||||
Db 1857 GTACCAAAAGACTTCATCACCCAAATACACAGAGAGCGTTTCGCTCAGAGAAGTAT 1916 |||||:|||||
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560 |||||:|||||
Db 1917 GCAGATAACGAGCTCGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1976 |||||:|||||
Qy 561 GlyLeuLeuGlnAlaTyRAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580 |||||:|||||
Db 1977 TCCCTGATCCAGGTCTATGCTGAAGACTGGACCTGATGGAGCACTTAATCAGGCTAAC 2036 |||||:|||||
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600 |||||:|||||

Db 2037 CAACATGAACCCAGCGGAGACATCACTACATCTTGGGTACGAATGGTGGACCGAAACTCC 2096 |||||:|||||
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGluThrClyLys 620 |||||:|||||
Db 2097 CTCCATATTGGGACTTTCTTGTAGAACAGTAGTGGCAATTTAGACAAGCAGACAGCCAAA 2156 |||||:|||||
Qy 621 GlySerThrAlaLeuHisIleTyRAspLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640 |||||:|||||
Db 2157 GGAAGCAGCGCTACATATTGCTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 2216 |||||:|||||
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660 |||||:|||||
Db 2217 CTGGGGGAAAGCATCTGTGAGCATTAATGATGATGATGATGATGATGATGATGATGATGAT 2276 |||||:|||||
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680 |||||:|||||
Db 2277 GCGCAGCGCTCAAACTCCAAATGCGAGGAGCTGTGACTCAGGCGCAGCAGCGGGAAG 2336 |||||:|||||
Qy 681 PheAsnSerHisValHisValGluTrpArgLeuLeuHisGluAspLeuAspGlu 700 |||||:|||||
Db 2337 TTCAATGTCATGTGCATGTGGAATATGACTGGCGCTGTCATATAGAGATCTGGACGAG 2396 |||||:|||||
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720 |||||:|||||
Db 2397 AGCAAGATGATGATGAGGACAAG-----CCCATTCATCAGCGGTGAGGAGCGTCCA 2450 |||||:|||||
Qy 721 IleSerPheTyRLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740 |||||:|||||
Db 2451 ATAAAGCTGTATAGTTCAGGCGAGTGGCCCCATGATGCCAACATGAGCGCTCTGGCTCGG 2510 |||||:|||||
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760 |||||:|||||
Db 2511 GACGTGGCCATGTGGTCAATAATAAGCAGAGGCGCTTTATTCGAGGATGATGATGAAC 2570 |||||:|||||
Qy 761 GluThrTyRArgAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780 |||||:|||||
Db 2571 GAGACTTACGGCAGCATGCTCGATCCCACTCTCCACCACCTG-----GGTTTACCAGGA 2624 |||||:|||||
Qy 781 ThrThrSerAlaProProLeuProArgAsnValGlyLysValGlnThrAlaSerSer 800 |||||:|||||
Db 2625 GTACTGGCATCTCTTTTACCCTCGGCGCTTGGGAAGGGGA----- 2669 |||||:|||||
Qy 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820 |||||:|||||
Db 2670 -----TGG-----AGTCCACCAATGGAGAACATCGGTAGACAGAGGTCA 2708 |||||:|||||
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840 |||||:|||||
Db 2709 TGTTCAGATCCTGCA-----AACCTCAAACTCTCGAACAAATAAATCTCTGTGTAT 2759 |||||:|||||
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860 |||||:|||||
Db 2760 GTTCTGCTCTGCTCTCTCCACCTCTCTCTGCA----- 2792 |||||:|||||
Qy 861 SerGlnProSerLysProAlaProGlyIleSerGlnIleArgProProProLeuPro 880 |||||:|||||
Db 2793 -----CCCAAGAGACCTCCACCTCCAGATCCAAAGGCCAGTCTTCTCTCCACAGCC 2846 |||||:|||||
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900 |||||:|||||
Db 2847 ACGGCTCCT-----CCTGCACCA-----TCCGCA 2870 |||||:|||||
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920 |||||:|||||
Db 2871 CCGTCTCTTATTCACACCTGCTCTCTCAGGCGAGCGCT----- 2909 |||||:|||||
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940 |||||:|||||
Db 2910 -----GTAGTGGCCCTCGACCACTGATGATGATGATGATGATGATGATGATGAT 2945 |||||:|||||
Qy 941 AlaThrLysLeuLysProLysArgValLysAla 951 |||||:|||||
Db 2946 CTGACTGATGTCAAAGGTCTGCTGTCTAAAGCC 2978 |||||:|||||

```

RESULT 12
BC048818
LOCUS
DEFINITION
Mus musculus development and differentiation enhancing, mRNA (CDNA
clone IMAGE:3668778), partial cds.
ACCESSION
BC048818
VERSION
BC048818.1 GI:28981428
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4434)
Strausberg,R.I., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L.H., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Udwin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Rulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butcherfield,Y.S., Krzywinski,M.I., Skalek,U., Smallos,D.E.,
Schnur,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4434)
Strausberg,R.
Direct Submission
Submitted (14-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nigri.nih.gov
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 99 Row: a Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1. .4434
/organism="Mus musculus"
/mol_type="mRNA"
/strain="mix FVB/N, C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3668778"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI CGAP_Mam5"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
<1. .4434
/gene="Ddef1"
/notes="synonyms: PAP, a19"
/db_xref="LocusID:13196"
/db_xref="MGI:1342335"
<1. .3504
/gene="Ddef1"
/codon_start=1
/product="Ddef1 protein"
/protein_id="AAH48818.1"
/db_xref="GI:28981429"
/db_xref="LocusID:13196"
/db_xref="MGI:1342335"
/tranlation="APLAPLASAPLLARQAGRRGGCGARGVHEPPGPGSGRRDRSRP
PPLVSRSGRMPDQISVSEFIATETEDYNSPTSSFTRLHNCENTVTLLEALDQD
KELSTLKNLQGLSHNVFTDLSLLGDLKVGDKLKKPFDKAWDQYETRFKLEKE
RTALQKVKSKAIYNSQDHFVQNEENYAQVLDKFGSNFLSRDNPDLGTAFKFKTLT
KREHAKGHMIRTEITGAEIAEEMKERRFLQMCQCEYLKIVNEIKTKGDLQNLQI
KYHAQCNFFQDGLTKADKLQYIEKLAADLYNIKOTQDEBKQLTALRDLTKSSLOI
DKESOSRQGYSMHOLQNGEYKSGKFLKKSGDGRKVMORRCAKNGILTIS
HATSNOPAKNLNTLQKVNADKSKFDLISHNRTVHFQAEQDYATMISVLTNSK
BEALTMFRGQSQTGENSLELTALIEDVORLPNDICCGSGSEPTWLTNLGILT
CIECSGIRHMGVSHISLELDKLTSELLAKNVGNNSFNDIMEANLSPSPKPT
PSPMTVRKEYITARYVDHFRSRTKTCASSAKJNELLEAIKSRDLALIQVYAEVGL
MEPLCEQELGETALHVAQTQSLHLDVFLVQCNLDKQTSVGNTVLHVSGMY
KPECLKLLRSKPTVDIVNONGTALDIKRLKATOCEDLLLSQAKSKFNPHVVEY
EWNLRQDEMDESDDLDDKPSIKKESPRQSSCHSSISPODKLAPGSPTRDKQ
RLSGAFNTQIFASTDSTLPSPTSEAPLPFRNKGKTPPSTLPLGTQTSSGSSST
LSKKRPPPPPPGKHRTSLDPSPLPHGPNKAIPWGNVGLPSSSTANKPEGUSQQ
ASTSATAKALGRVLPKLQPKVALRKTETSHLLDRTNIPETFKSSQSLTELPPK
PLGLPLPKPVELAPKQVGLPPKPGELPPKQLDLPKQLDLPKQSPKQMDLPPK
POLGDLAKSOAGVSAKVQPPSVTQSHGSDLSNVQSRDAIKQASESDNDLPTT
LPEPTVPLPKINTGKNKVRVKTIYDCAQNDDELTFIEGEVIVTGEDEWIGH
TEGQPERKGVFPVSVFVHLSD"
1084. .1359
/gene="Ddef1"
/notes="PH; Region: Pleckstrin homology domain. Domain
commonly found in eukaryotic signalling proteins. The
domain family possesses multiple functions including the
abilities to bind inositol phosphates, and various
proteins. PH domains have been found to possess inserted
domains (such as in PLC gamma, sytrophins) and to be
inserted within other domains. Mutations in Brutons
tyrosine kinase (Btk) within its PH domain cause X-linked
agammaglobulinaemia (XLA) in patients. Point mutations
cluster into the positively charged end of the molecule
around the predicted binding site for phosphatidylinositol
lipids"
/db_xref="CDD:smart00233"
1420. 1761
/gene="Ddef1"
/notes="ArfGap; Region: Putative GTPase activating protein
for Arf. Putative zinc fingers with GTPase activating
proteins (GAPs) towards the small GTPase, Arf. The GAP of
ARD1 stimulates GTPase hydrolysis for ARD1 but not ARFb"
/db_xref="CDD:pfam01412"
1912. .2172
/gene="Ddef1"
/notes="ANK; Region: ankyrin repeats"
/db_xref="CDD:cd00204"
3325. 3492
/gene="Ddef1"
/notes="SH3; Region: Src homology 3 domains"
/db_xref="CDD:cd00174"

```

ORIGIN

Alignment Scores:			
Pred. No.:	5.48e-104	Length:	4434
Score:	3018.00	Matches:	621
Percent Similarity:	65.88%	Conservative:	128
Best Local Similarity:	54.62%	Mismatches:	232
Query Match:	57.23%	Indels:	156
DB:	10	Gaps:	18
US-09-914-042-1 (1-1006) x BC048818 (1-4434)			
Qy	1	MetProAspGlnIleSerValserGluPheValalGluThrHisGluAspTyrLysAla	20
Db	166	ATGCCGACCAAGATCTCGTGTCGAGTTCACTCCCGAGACCACCGAGGACTACAACTCG	225
Qy	21	ProThrAlaserSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	226	CCCACCAGTCCAGCTTCACCACCGCGGTGCACAACATGCAGGAACACCGTCACCGTCGTG	285
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla	60
Db	286	GAGAGAGCTCTAGACCAAGATAGAACACGCCCTACAGAAAGTGAAGAAATCTGTAAAAACA	345
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu	80
Db	346	ATATATAATCCGGTCAAGATCATGTACAAAATGAAGAAAACACTATGCACAAGTCTCGAT	405
Qy	81	LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys	100
Db	406	AAGTTTGGAACTAATTTTTTAAGCAGAGACAACCTGACCTTGACACCGCTTTTGTCAAG	465
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	466	TTTTTCTACATCACAAAGGAACCTGTCCACGCTGCTGAAAAATCTGCTCCAGGGTCTGAGC	525
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	526	CACAAATGTGATCTTCACCTTGATTCCTTGTATAAGAGAGACTTGAAGGGGTCAAAGGA	585
Qy	141	AspLeuLysLysProPheAspLysAlaTPrLysAspTyrGluThrLysIleThrLysIle	160
Db	586	GATCTCAAGAAGCCATTGTGACAAAGCCCTGGAAAGATTATGAGACGAAGTTTACAAAAATT	645
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
Db	646	GAGAAGGAGAAAGAGAGACGCGAAAAACAGCACGGGATGATCCGACAGAGATAACAGGC	705
Qy	181	AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu	200
Db	706	GCTCAGATTGCAGAAGAAATGGAAGAAAGGAACCGCGGCTTTTTCAGCTCCAGATGTGTGAA	765
Qy	201	TyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	766	TATCTCTATAAAGTAAATGAAATCAAGACCAAAAAGGGGTGGACCTGCTCAGAACCTT	825
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	826	ATAAAGTATTACCATGACATGCAATTTCTTTTCAGGATGGCTTGAAGAACGGCTGATAAA	885
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	886	CTGAAACAGTACATAGAAAAAGCTGGCTGCATTATATAATATAAGCAAAACCCAGAT	945
Qy	261	GluGluArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	946	GAAGAAAAAACAAGCTAACCTGCACTCCGAGATCTATAATAATCTCTCTGCAACTTGAT	1005
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1006	CCGAAAGAAAGACTCTCAGAGCCGGCAAGGT---GGGTACAGCATGCAATCAGCTCAGGGC	1062
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320

LAKVGNNSFNDIMEANLPSPPKPTSPSSDMTVERKEYITAKYVVDHRSRKTCSSSSAK
 LNELLEATKSDLLALIOVYAAEGVELMEPLLEPGELGETALHLAVRADTOTSLSHLVD
 FLVQNCNGLDKOTALGNLTHALVCSYKPECLKULLRSKPTVDVNVQAGETALDIAR
 LKATQCEBLLSQAKSKCNPHVVEYNLROEMDESDDLLDKPSIKKERSPRPQ
 SFCHSSLSIPQDKLSLPGFSTRDRKQSLSYGAFNQIFVSTSDSPSIAEAPPLPP
 RNATGPPSPSTPLSTQTSSGSTLSKRSPPPPGPKHRTLSDDPPSPPLPHGPNKG
 AVPMGNDVGPSSSKXTNKFGLSQSSQTSKALGPRLPKLPQKVALKRTTSHH
 LSLDKANVPPIEFOKSSOLTELPOKPPPGDLPKPPTELAPKPPITGLDLPKPGELPPKP
 QLGLDLPKPEQLADLPPKQVQDLPPKPOLGELLAKPQTGDASPKPAQPPLELTPKSHPA
 DLSNPVKQASEDINDLTPETPLPRKNTGSKVRKTYIDCQANDDELTF
 MEGEVIVVTGEEDEQEWIGHIEGQPERKGVFPVSVHLSQ

ORIGIN

Alignment Scores:
 Pred. No.: 9,77e-104 Length: 5330
 Score: 3013.50 Matches: 627
 Percent Similarity: 66.67% Conservative: 129
 Best Local Similarity: 55.29% Mismatches: 225
 Query Match: 57.15% Indels: 153
 Gaps: 23

US-09-914-042-1 (1-1006) x AF112886 (1-5330)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20
 Db 269 ATGCCGGACCATGCTCCGTCCTCCGAGTTCATCGCGGAGACCACCGAGGACTACAACTCG 328
 Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 Db 329 CCCACCGCTCCAGCTTCCTACTACGGGGCTGCACAACTGCAGGAACACCGTCACTGCTGCT 388
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrIysMetLysLysSerValIysAla 60
 Db 389 GAGGAGGCTCTAGACAAGATAGAACAGCCTTACAGAAAGTTAAGAGTCTGTAAAGCA 448
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
 Db 449 ATATACAAATTCGGTCAAGACCATGTACAAATGAAGAAACTATGCCGCAAGTTCTTGAT 508
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
 Db 509 AAGTTTGGGAGTAATTTTTTAAGTCGAGACACCCAGATCTTTGGCACCCTTTTGTCAAG 568
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
 Db 569 TTTTCTACGCTTACAAAGAACTGTCCACACTGCTGAAATAATCTGCTCCAGGGCTGAGC 628
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
 Db 629 CACAATGTGATCTTCACCTTGGATTCCTTTGTTGAAGGAGAGCTGAAGGAGTCAAGGC 688
 Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
 Db 689 GATCTCAAGAAACCAATTTGACAAAGCTTGGAAAGATTATGAGCAAGGTTTACCAAAAT 748
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
 Db 749 GAGAAGGAGAGAGGAGGAGCACCAGCAGCAGGAGATGATCCGACGAGATACCCGC 808
 Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
 Db 809 GCCGAGATCGCGAGGAAATGAAAGAGGAGCGCGCTCTTCCAGCTCCAGATGTGCGAG 868
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
 Db 869 TATCTCAATTAAGTTAATGAATCAAGACCAAAAGGGGTGTGGATCTGCTGCAGAACCTG 928
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
 Db 929 ATAAAGTATTATCACGCACAGTGCATTTCTTTCAAGATGGTTTGAAGACAGCTGATAA 988
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260

Db 989 TTGAAACAGTACATTGAAAGCTGGCTGCTGATTGTGTATATATCAAAACAGACCCAGCAG 1048
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGluValGlu 280
 Db 1049 GAAGAAAGAAACAGCTGACCGCACTCCGAGACCTAATAAAGTCTCGCTCCAACCTCAT 1108
 Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297
 Db 1109 CAGAAGGAGTCTAGAGAGATTCACAGACCGCGGAGGGA---GGCTACAGGATGCACCCAG 1165
 Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
 Db 1166 CTGCAGGCGCAACAAGGAATACCGCAGAGAAAGGCTACCTGCTGAGAAAGAGTGCAC 1225
 Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
 Db 1226 GGGATCCGGAAGGTGCGCAGAGAAGAGTCTCCGCTCAAGAACGGGATCCTGACCAATC 1285
 Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
 Db 1286 TCCACGCCAGCTCCACAGACAGCCAGCAAGCTGAACCTTCTCACTTGCAGCGTGAAG 1345
 Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe 377
 Db 1346 CCGAATCCGAGGACAAGAGTCTTTTACCTGATATACATAACAGAGCGTATACACTTT 1405
 Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
 Db 1406 CAGGCCAGAGATGACAGGATTTATGTACGCGGATTCGGTCTGCGACAAACAGCAAGAG 1465
 Qy 398 GluAlaLeuAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417
 Db 1466 GAGGCCCTCACCATTGGCTTCCGGGGGAAACAGAGTCTGGGGAGAGCAGCTG---GAG 1522
 Qy 418 GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
 Db 1523 GAGCTGACGAAGGCATCATCGAGAGCTGCAGCGGCTCCCGGGCAACAGAGCTGTGCTGC 1582
 Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457
 Db 1583 GACTGCGGCTCGGCAGAACCCACCTGCTGTCACCACTTGGGCATCTTGACCTGTATA 1642
 Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477
 Db 1643 GAATGTTCCGCATCCATAGAAATGGGGTTTCATATTTCTCGCATCCAGCTCTTGGAA 1702
 Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497
 Db 1703 CTAGACAAATTAGGAACCTCTGAACTCTTGTGCGCAAGAAATGTAGGAAACAATAGTTT 1762
 Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517
 Db 1763 AATGATATTATGGAAGCAAAATTTTACCAGT---CCCTCACCAAAACCCACCCCTTCAAGT 1819
 Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysIleGluArgArgTyrAlaArg 537
 Db 1820 GATATGCTGACGGAAGGAATATACCTGCAAGATATGTAGATCATAGATTTCACGG 1879
 Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557
 Db 1880 AAGACCTGTTTCATCGTCACTAGCTAAACTGAAGTTCCTTGTAGGCCATCAAAATCCAGG 1939
 Qy 558 AspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577
 Db 1940 GATTACTTGCATTAATTAAGTCTATGACAGGGGGTGGAGCTAATGGAACCGCTG--- 1996
 Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597
 Db 1997 CTGGNACCCGACAGGAGCTTGGGAGACACCCCTTTCATCTTGCAGTCCGAACCCGACAG 2056
 Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617
 Db 2057 CAGACATCTCTCCATTTGGTGGACTTCTTCTGTACAAAACCTGTGGGAACCTAGATAAGCAG 2116

Db 389 GAGGAGGCTCTAGACCAAGATAGAACAGCCTTACAGAAAGTTAAGAAAGTCTGTAAGACGA 448
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInTyThrGlnAlaLeuGlu 80
Db 449 ATATACAAATCCGGTCAAGACCATGTACAAATGAAGAAACTATGCCCAAGTTCTTGAT 508
Qy 81 LysPheGlyGlyAsnCyValCyArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 509 AAGTTTGGAGTAATTTTAAAGTCGACAGCAACCCAGATCTTGGCCCGCTTTGTCAAG 568
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 569 TTTTCTACGCTTACAAAGAACTGTCCACACTGTCTGAAATAATCTGCTCCAGGGCTGAGC 628
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 629 CACAATGTGATCTTCACCTTGATCTCTTGTGAAAGGAGAGCCTGAAAGGAGTCAAAGGC 688
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 689 GATCTCAAGAAACCAATTTGACAAAGCTTGGAAAGATTATGACAGAGTTTACCAAAATT 748
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 749 GAGAAAGGAGAGAGGGAGCAGCCCAAGCAGCAGCGGATGATCCGCAAGGAGATCACCGGC 808
Qy 181 AlaGluIleAlaGluLysGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 809 GCCGAGATCGCGGAGGAAATGAAAGAGGAGCGGCCCTCTTCCAGCTCCAGATGTGCGAG 868
Qy 201 TyrLeuLeuLysValAsnGluLysLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 869 TATCTCATTAAGTTAATGAATCAAGACCAAAAGGGTGTGGATCTCTCGCAGAACCTG 928
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 929 ATAAAGTATTATCACGCAGTGCATTTCTTCAAGATGGTTTGAAAGACAGCTGATAAA 988
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 989 TTGAACACAGTACATTGAAAGCTGCTGCTGATTTGTATATATATCAACACAGACCCAGAC 1048
Qy 261 GluGluArgArgGlnLeuGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1049 GAAGAAAGAAACACAGCTACGCCACTCCGAGACCTTAATGAAGTCTCTCTCACTCGAT 1108
Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297
Db 1109 CAGAAGGAGTCTAGGAGAGATTCCACAGCGCGGAGGGA---GGCTACAGCATGCACCAG 1165
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
Db 1166 CTGAGGCGCAACAGGAATACGGCAGCAGAGAAAGGGCTACTCTCTGAAAGAGTAGTAC 1225
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 1226 GGGATCCCGGAAGTGTGCGAGAGAAGAGTGTCTCCGTCAAGACCGGATCTCTGACCATC 1285
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuLeuThrCysGlnValLys 357
Db 1286 TCCCAAGCCACGCTCCCAACAGACAGCCAGCTGAACCTTCTCACTTGCAGGTGAAG 1345
Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe 377
Db 1346 CCGAATGCCGAGGACAAGAAGTCTTTTGACCTGATATCACATAACAGCAGCTATCACTTT 1405
Qy 378 GlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
Db 1406 CAGGCCGAAGATGACAGAGTATTAGCTGGATCTCTGGTGCTGACAAACAGCAAGAG 1465
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGln 417
Db 1466 GAGGCCCTCACCATGGCTTCGCGGGGGAACAGAGTGTGTGGGAGAGCAGCTG---GAG 1522

Qy 418 GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
Db 1523 GAGCTGACGAAGGCCATCATCGAGAGGTGCGGGCTCCCGGCAACAGAGTCTGCTGC 1582
Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457
Db 1583 GACTTCGGCTCGGCAGAACCCACCTGGCTGCCACCACTTGGGCACTTTCGACCTGTATA 1642
Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477
Db 1643 GAATGTTTCGGCATCCATAGAGAAATGGGGTTCATATTTCTCGCATCCAGTCTTTGAA 1702
Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497
Db 1703 CTAGACAAATAGGAACCTTCTGAACTCTTGTGGCAAGATGTAGGAAACAAATAGTTT 1762
Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517
Db 1763 AATGATATTGGAAGCAAAATTTACCCAGT---CCCTCACCACCAACCCACCCCTCAAGT 1819
Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537
Db 1820 GATATGACTGTACGGAAGGAATATATCACTGCAAGATGTAGATCATAGTGTTCACGG 1879
Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557
Db 1880 AAGACCTGTTTCATCGTCACTAGCTAAACTGAACGAATGCTTGGAGGCATCAAAATCCAGG 1939
Qy 558 AspilePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577
Db 1940 GATTACTTGCACATAATTCAGTCTATGCAGAGGGGTGGAGCTAATGGAACCCGTG--- 1996
Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597
Db 1997 CTGGAACCCGACAGGAGCTTGGGGAGACAGCCCTTCATCTTGCAGTCCGAACCCGACAG 2056
Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617
Db 2057 CAGACATCTCTCCCAATTTGGTGGACTTCTTGACAAACCTGTGGAACTAGATAGGAG 2116
Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeu 637
Db 2117 ACGGCCCTGGGGAACACGGCCCTGCACCTACTGTAGTATGTACAGTAAACACAGAGTGTG 2176
Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657
Db 2177 AAGCTGTCTCTCAGGAGCAGCCCACTGTGACCTGCTTAATCAGGCTGGAGACCGCC 2236
Qy 658 LeuAspileAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677
Db 2237 CTGACATAGCAAGAGACTGAAAGCCACTCAGTGTGAAGACCTGTCTTCCAGCTAAA 2296
Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697
Db 2297 TCTGGAAGTTCAATCCACAGCTCCACCTGGAATATGATGGAATCTTCGACAGGAGAG 2356
Qy 698 LeuAspGluSerAspAspMetAspMetAspGluLysLeuGlnProSerGluAsnArgArgGlu 717
Db 2357 ATGATGAGAGCGATGACGACTGGATGACAAACCGAGCCCGCCCATCAAGAGAGCGCTCC 2416
Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737
Db 2417 CCCCAGCCGAGAGCTTCTGCACCTCTCCAGCATCTCCCCCAGGAC-----AAG 2467
Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGlnArgAlaPheMetProSerIle 757
Db 2468 CTCTCAGCTCCGGGCTTCAGCACCGCAAGGCAAGCAACGACTC----- 2512
Qy 758 LeuGlnAsnGluThrTyrGlyAla-----LeuLeuSerGlySerPro 771
Db 2513 -----TCTACGGCGCCTTCACCAACACAGATCTTCGTCTCCACAGCACAGAC 2560


```

Db      |||||ATCGCGGACAGATCTCGGTGTCGAGTTTCATCGCGGAGACCACCGAGGACTCAACACTCG 169      :
Qy      |||||ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40      :
Db      |||||CCCAACCAAGTCCAGCTTCCACACGCGGTGCACAACTGCAGAGAACACCGTCCACGCTGCTG 229      :
Qy      41  GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIleAla 60      :
Db      |||||GAGGAGGCTCTAGACCAAGATAGAACACGCCCTACAGAAAAGTGAAGAAATCTGTAAGACA 289      :
Qy      61  IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80      :
Db      |||||ATATAATTCGGTCAAGATCATGTACAAATGAAGAAACTATGCACAGTCTTGAT 349      :
Qy      81  LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100      :
Db      |||||AAGTTTGAAGTAATTTTAAAGCAGACAAACCTGACCTTGGCAGCGCTTTTGTCAAG 409      :
Qy      101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120      :
Db      |||||TTTTCTACACTCAAAAGGAATGTCACGCTGCTGAAAAATCTGCTCCAGGGCTTGAGC 469      :
Qy      121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140      :
Db      |||||CACAAATGTGATCTTCACCTTGATCTTGTAAAGGAGAGACCTGAAGGGGGTCAAGGA 529      :
Qy      141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160      :
Db      |||||GATCTCAAGAGGCCATTTGACAAAGCTTGAAGAGATTATGAGCGAAATTTACAAAAATT 589      :
Qy      161 GluLysGluLysLysGluHisAlaLysLysLeuHisGlyMetIleArgThrGluIleSerGly 180      :
Db      |||||GAGAGGAGAAAGAGAGACGCAAAACAGACCGGGATGATCCGAACAGAGATAACAGGC 649      :
Qy      181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200      :
Db      |||||CTGAGATTGAGAGAAATGGAAAAAGAACGCGGGCTTTTCAGCTCCAGATGTGTGAA 709      :
Qy      201 TyrLeuLeuLysValAsnGluIleLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220      :
Db      |||||TATCTCATTAAGTAATTAATCAAGACCAAAAGGGGTGTGGACCTGCTGCAGAACCTT 769      :
Qy      221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyValLeuAlaValGluSer 240      :
Db      |||||ATAAAGTATTACCATGCACAGTGAATTTCTTTCAGGATGGCTTGAAAACCGGCTGATAAA 829      :
Qy      241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260      :
Db      |||||CTGAACAGTACATAGAAAAGCTGCTGCTGACTTATATATATATAAGCAACCAACCAAGAT 889      :
Qy      261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280      :
Db      |||||GAAGAAAAAAGACAGCTAACTCCGACACTCCGAGATCTTAATAAAATCCTCTCTGCAACTGAT 949      :
Qy      281 GlnLysGlu-----AspSer 285      :
Db      |||||CCGAAGAAGTAGTGTTTATATGTTGCCAGCAGGGCTAAACAGTTCTAGAAGAGACTCT 1009      :
Qy      286 GlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlnLysLysGluHisGly 305      :
Db      |||||CAGACCCGGCAAGGT---GGGTACAGCATGATCATGCTCCAGGGCAACAGAGGATACGGC 1066      :
Qy      306 ThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLys 325      :
Db      |||||AGTGAAGAAGAGGGCTTCTGCTTAAAGAGAGTGTGGATCCGGAAGTGTGGCAGAGA 1126      :
Qy      326 ArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgPro 345      :
Db      |||||CGGAAGTGTGGGTCAAGAACGGGATCTGTACCATCTCCACGCAACTTCCAACAGGGCAG 1186      :
Qy      346 ProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCys 365      :

```

```

Db      1187 CTTGCTAAGTTAAACCTCTCTCACCTGCTCAAGTGAACCCGAATGCTGAGGACAAGAGTCT 1246      :
Qy      366 PheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCys 385      :
Db      1247 TTTGACCTGATATACATAATCGAACAATATACATTTCAAGCAGAAGATGACGAGATTAT 1306      :
Qy      386 GlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAlaPheLys 405      :
Db      1307 ATAGCGTGGATATACGTACTGACGAATAGCNAAGAGAGGCCCTAACCATGGCTTCCGT 1366      :
Qy      406 GlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleSer 425      :
Db      1367 GGTGAGCAGACGACAGCGGAGAAATAGCTG---GAGGACCTTACCAAGGCCATCATCGAG 1423      :
Qy      426 GluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThr 445      :
Db      1424 GATGTACAGCGGCTCCCTGGGAATGACATCTGCTGTGACTCGCGCTCATCAGAACCCACG 1483      :
Qy      446 TrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGlu 465      :
Db      1484 TGGCTTTCAACCAACTTGGGTATTTTGCCTGTATAGATGTTCCGGAATCCATAGGAA 1543      :
Qy      466 LeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGlu 485      :
Db      1544 ATGGGGTTTCATATTTCTCGCATTCAGTCTTTGAACTAGACAAATTTAGGACTTCGAA 1603      :
Qy      486 LeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysLeu 505      :
Db      1604 CTCTTGTGTCGCAAGATGTAGGAACAATAGTTTAAATGATATTTATGGAAGCAAAATTTA 1663      :
Qy      506 ProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyr 525      :
Db      1664 CCCAGC---CCTTCACCAAAAGCCACCCCTTCAGTGTATGACCGTGCCTTAAGAGTAC 1720      :
Qy      526 IleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAla 545      :
Db      1721 ATCACTCAAAAGTATGTCGACCACCGCTTCTCCAGGAAGACCTGTGCTCTCTCCACGA 1780      :
Qy      546 LysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAla 565      :
Db      1781 AAGCTGAACCAATTTGCTTGAGGCCATCAAGTCCAGGAGCTTACTTGGATTAAATCCAAGTC 1840      :
Qy      566 TyrAlaAspGlyValAspLeuThrGlyLysIleProLeuAlaAsnGlyHisGluProAsp 585      :
Db      1841 TATCAGAGGGGTGGATTAATGAGCCACTG---CTGGNAACCTGGCGAGGACTCGGG 1897      :
Qy      586 GluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAsp 605      :
Db      1898 GAGACAGCCCTTCATCTTGGCTCAGGACCGCAGACAGACATCTCTCCATCTGTTGAT 1957      :
Qy      606 PheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeu 625      :
Db      1958 TTCCTTGTACAAAACCTGTGGGAACCTGATAAGCAACGCTGTGGGGAATACAGTTCTG 2017      :
Qy      626 HisTyrCysCysLeuThrAspAsnAlaGlyCysLeuLysLeuLeuLeuArgGlyLysAla 645      :
Db      2018 CACTATTGACAGCTGTACGGCAAGCCGAGTGTCTGAAGCTGCTGTAGGAGCAAGCCCC 2077      :
Qy      646 SerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLys 665      :
Db      2078 ACCGTGACATCTGTTAACCAAGATGGAGAACTGCCCTGGATATAGCAAGAGACTCAAA 2137      :
Qy      666 HisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisVal 685      :
Db      2138 GCTACCCAGGTGTGAAGATCTCTTCCAGGCTAAATCTGGAAAGTTCAACCCATCATGTC 2197      :
Qy      686 HisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGluSerAspAspMet 705      :
Db      2198 CACGTAGATGATGATGGAATCTTCGACAGACAGATGGATGAGAGCGATGACGATCTG 2257      :
Qy      706 AspGluLysLeuGlnProSerGluAsnArgGluAspArgProIleSerPheTyrGln 725      :
Db      2258 GATGACAAAGCCGCGATCAAAAAGGAACGCTCGCCACAGACCACAGAGCTTCTGCCAC 2317      :

```

Search completed: August 4, 2005, 11:54:31
Job time : 10750 secs

```
Qy 726 LeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAsnLeu 745
Db |||||
2318 TCCTCCAGCATCTCGCCACAGAC-----AAGCTGGCACTGCCAGGTTTCAGCACT 2368
Qy 746 AlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAla 765
Db |||||
2369 CCACGGGACAAAGCAGCGCTC-----TCCTACGGAGCC 2401
Qy 766 Leu-----LeuSerGlySerProProAlaGlnProAlaAlaProSerThrThr 782
Db |||||
2402 TTCACCAACCAAGATCTTCGCCTCTACGAGCAGACAGATTGGCCACATCACCC---ACCAGT 2458
Qy 783 SerAlaProProLeuProProArgAsnValGlyLys-----ValGlnThrAlaSer 799
Db |||||
2459 GAGGCTCCCTTTGTCACCTCGGAACCGCCGGAAAGGTAATGATGTGGGCCCATCATCT 2518
Qy 800 SerAlaAsnThrLeuThrLysThrAsnSerValSerValAspGlyGlySerArgGlnArg 819
Db |||||
2519 TCAAGTAAGCGCAAAACAAGTTTGAGGGGTGTCTCAGCAAGCAAGCACCAGTCTTGCT 2578
Qy 820 SerSerSerAspProProAlaValHisProLeuPro-----ProLeuArgVal 836
Db |||||
2579 AAGACTGCCCTTGGCCCGAGAGTCTTCTAACTACCTCAGAAAGTGGCACTAAGGAAG 2638
Qy 837 ThrSerThrAsn----- 840
Db |||||
2639 ACGGAGACCAAGCCATCATCTCTCCCTCGCAGAACCAACATCCACCTCGAGACTTTTCAG 2698
Qy 841 -----ProLeuThrProThrProProPro 848
Db |||||
2699 AAATCATCACAGTTGACAGATTACCCCAAAAGCCACCACCTTGGAGAGCTGCCCCGGAAG 2758
Qy 849 ProVal-----AlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys--- 865
Db |||||
2759 CCTGTGGNACTGGCCCCCAAGCCCAAGTTGGAGAGCTGCCACCTAAGCCTGGAGAGCTA 2818
Qy 866 -----ProAlaProProGlyIleSerGlnIleArgPro 876
Db |||||
2819 CCCCTAAGCCCCAATTAGGTGACCTGCCCCCAAGCCACAGCTCTCAGACTTACCTCCC 2878
Qy 877 ProPro-----LeuProPro----- 881
Db |||||
2879 AAGCCACACAGATGAAGACCTGCCCCCTAAGCCGAGCTGGGGGATCTGCTGGCAAGTCC 2938
Qy 882 -----GlnProProSerArgLeuProGlnLysLysPro 892
Db |||||
2939 CAGGCTGCGGATGTCTCAGCCAAGGTGCGCCACCCCTCAGAGGTACACACAGG----- 2992
Qy 893 AlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProVal 912
Db |||||
2993 -----TCACACACCGGGATCTGTCTCCAAATGTACAGTCCAGAGATGCCATC 3040
Qy 913 AspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProPro 932
Db |||||
3041 CAGAAGCAAGCATCTGAAGACTTCCAAAGCAGCTCACACCCACGCTGCCA---GAGACACCC 3097
Qy 933 AlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeu 952
Db |||||
3098 GTACCACTGCCGAGAAAATCAATACGGGGGAAATAAATGAGCGCGGTGAAGACCAATT 3157
Qy 953 TyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIle 972
Db |||||
3158 TATGACTCCAGGCAGATAATGACGACCACTCATATTATTGAGGGGGAAGTGTATCAT 3217
Qy 973 ValAspGlyGluGluAspGlnGluThrTrpIleGlyHisIleAspGlyAspProGlyArg 992
Db |||||
3218 GTCACCGGGGAAGAGGACAGGAGTGTGGATCGGGCATATCGAAGGACAGCCTCAAGG 3277
Qy 993 LysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db |||||
3278 AAGGTGTCTTTCAGATGTCTTTTGTCCACATCTCTGTCTGAC 3319
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 4, 2005, 08:26:55 ; Search time 103 Seconds
(without alignments)
3813.012 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAFVPSVHFIAID 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1752860 seqs, 390397842 residues

Total number of hits satisfying chosen parameters: 1752860

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3006	57.0	1132	15	US-10-210-281-74
2	2071.5	39.3	903	15	Sequence 74, Appl
3	2071.5	39.3	903	17	Sequence 2951, Ap
4	1801.5	34.2	303	11	Sequence 2, Appli
5	1759.5	33.4	349	11	Sequence 635, App
6	1075.5	20.4	208	15	Sequence 956, App
7	611.5	11.6	358	16	Sequence 2284, Ap
8	569.5	10.8	778	15	Sequence 2815, Ap
9	558	10.6	764	15	US-10-080-334-238
10	551	10.4	804	15	Sequence 238, Appl
11	533	10.1	759	15	Sequence 80, Appl
					Sequence 235, App
					Sequence 236, App

12	529	10.0	834	14	US-10-176-306-11	Sequence 11, Appl
13	529	10.0	834	15	US-10-080-334-237	Sequence 237, App
14	529	10.0	834	15	US-10-467-434-4	Sequence 4, Appli
15	509.5	9.7	775	16	US-10-739-930-6184	Sequence 6184, Ap
16	509	9.7	740	15	US-10-080-334-239	Sequence 239, App
17	487.5	9.2	776	16	US-10-729-930-6289	Sequence 6289, Ap
18	480	9.1	156	15	US-10-276-774-1638	Sequence 1638, Ap
19	438.5	8.3	773	16	US-10-437-963-136533	Sequence 136533,
20	419	7.9	605	9	US-09-828-303-19	Sequence 19, Appl
21	419	7.9	605	16	US-10-716-089-19	Sequence 19, Appl
22	408.5	7.7	503	11	US-09-833-245-1843	Sequence 1843, Ap
23	406	7.7	136	15	US-10-276-774-1754	Sequence 1754, Ap
24	398	7.5	324	17	US-10-921-707-3	Sequence 3, Appli
25	396.5	7.5	732	16	US-10-437-963-181247	Sequence 181247,
26	391.5	7.4	649	16	US-10-437-963-181939	Sequence 181939,
27	340.5	6.5	882	15	US-10-334-143-9	Sequence 9, Appli
28	330.5	6.3	369	16	US-10-425-115-233034	Sequence 233034,
29	330.5	6.3	580	15	US-10-094-749-2012	Sequence 2012, Ap
30	323.5	6.1	726	15	US-10-467-434-17	Sequence 17, Appl
31	321	6.1	440	15	US-10-424-599-160260	Sequence 160260,
32	311	5.9	336	15	US-10-424-599-158918	Sequence 158918,
33	306	5.8	467	15	US-10-108-260A-3678	Sequence 3678, Ap
34	296	5.6	836	16	US-10-719-993-842	Sequence 842, App
35	296	5.6	836	16	US-10-719-993-843	Sequence 843, App
36	296	5.6	836	16	US-10-719-993-844	Sequence 844, App
37	295	5.6	856	16	US-10-719-993-845	Sequence 845, App
38	292.5	5.5	886	15	US-10-321-625-6	Sequence 6, Appli
39	279	5.3	718	13	US-10-087-192-333	Sequence 333, App
40	278	5.3	802	16	US-10-684-422-246	Sequence 246, App
41	271	5.1	1205	15	US-10-332-416-1	Sequence 1, Appli
42	267.5	5.1	814	17	US-10-781-581-185	Sequence 185, App
43	265	5.0	383	16	US-10-425-115-326274	Sequence 326274,
44	262.5	5.0	319	16	US-10-425-115-225159	Sequence 225159,
45	261.5	5.0	348	15	US-10-425-114-64367	Sequence 64367, A

ALIGNMENTS

RESULT 1

US-10-210-281-74

; Sequence 74, Application US/10210281

; Publication No. US20040030096A1

; GENERAL INFORMATION:

; APPLICANT: Gorman, Linda

; APPLICANT: Zerhusen, Bryan D.

; APPLICANT: Edinger, Shlomit R.

; APPLICANT: Padigaru, Muralidhara

; APPLICANT: Guo, Xiaojia

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Zhong, Mei

; APPLICANT: Patutarajan, Meera

; APPLICANT: Miller, Charles E.

; APPLICANT: Ji, Weizhen

; APPLICANT: Pena, Carol E.A.

; APPLICANT: Burgess, Catherine E.

; APPLICANT: Sciore, Paul

; APPLICANT: Stone, David J.

; APPLICANT: Taupier, Raymond J., Jr.

; APPLICANT: Cagman, Stacie

; APPLICANT: Rothenberg, Mark E.

; APPLICANT: Malyankar, Uriel M.

; APPLICANT: Boldog, Ferenc L.

; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS

; TITLE OF INVENTION: THE SAME

; FILE REFERENCE: 21402-416D

; CURRENT APPLICATION NUMBER: US/10/210,281

; CURRENT FILING DATE: 2003-02-05

; PRIOR APPLICATION NUMBER: 60/309,501

; PRIOR FILING DATE: 2001-08-02

; PRIOR APPLICATION NUMBER: 60/310,291

; PRIOR FILING DATE: 2001-08-03

; PRIOR APPLICATION NUMBER: 60/361,775

```
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/361,832
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,203
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,201
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,702
; PRIOR FILING DATE: 2001-08-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 74
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-210-281-74

Query Match      57.0%; Score 3006; DB 15; Length 1132;
Best Local Similarity 55.3%; Pred. No. 6,8e-161;
Matches 626; Conservative 127; Mismatches 234; Indels 144; Gaps 22;

Qy 1 MPDQISVSEFAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMVLKMKSVKA 60
Db 21 MPDQISVSEFAETHEDYKAPTASFTTRTAQCRNTVAABEALDVRMVLKMKSVKA 80

Qy 61 INSSGLAHVNEBQYTOALEKFGNCVCRDDPDLGSAFLKFSVFTKELTALFKMLIQNMN 120
Db 81 IYNSGQDHVQNEENYAQVLDKFGSNFLSRDNPDLGTAFFKFTSTLKSLKLLKQLQGLS 140

Qy 121 NIISFPLDLSLKGDKLVKQKDPDKAWDYETKITKIEKKEHAKLHGMIRTEISG 180
Db 141 HNVIPTDLSLKGDKLVKQKDPDKAWDYETKITKIEKKEHAKLHGMIRTEISG 200

Qy 181 AEIAEEMEKERRFFOLQCEYLLKVEIKIKKGVLDLQNLKYPHAQCNFFQDGLKAYES 240
Db 201 AEIAEEMEKERRFFOLQCEYLLKVEIKIKKGVLDLQNLKYPHAQCNFFQDGLKATDK 260

Qy 241 LKPSIETLSTDLHTIKQADERRQLIQRLILKALQVQKE---DSQIRQSTAYSLSHQ 297
Db 261 LKQVIEKLAADLYNIKQVDEEKKQLTALRDLIKSLQDQKESRRDSQSRQG-GYSMHQ 319

Qy 298 PQGNKEHETERNGSLYKSDGIRKVKQKCSVKNGFLTISHGTANRPPAKNLILLTCQVK 357
Db 320 LQGNKEYSEKGYLLKSDGIRKVKQKCSVKNGFLTISHGTANRPPAKNLILLTCQVK 379

Qy 358 TNPEKKCFDLISHDRTHYFAEDRQEQCIQWMSVLQNSKEEALNNAFFKGDNDTGNNTIVQ 417
Db 380 PNAEDKKSFDLISHNRTHYFAEDRQEQDYVAMISVLNTSKEEALTWAFRGEQSAENSL-E 438

Qy 418 ELTKETIIEVQRMONTDVCDCGADPTWLTSTNLGILTCIECGSHRELGVHSPMQSLT 477
Db 439 DLTKAILEDVQRLPNDICCDGSEPTWLTSTNLGILTCIECGSHREMGVHISIQSLE 498

Qy 478 LDVLGTSLELLAKNIGNAGFNEIMCCCPAEDSVKPNPNSDMNARKDYITAKYIERRYAR 537
Db 499 LDKLGTSELLAKNIGNAGFNEIMCCCPAEDSVKPNPNSDMNARKDYITAKYIERRYAR 557

Qy 538 KKHADNAAKLSLCEAVKTRDIFGLQAYADGVLTETKIPIANGHEPDETALHLAVRSVD 597
Db 558 KTCSTSSAKNELLEFAIKSRDOLLALIOYVAGVELWEPL-LEPGQELGETALHLAVRTAD 616

Qy 598 RTSLSHIVFLVQNSNLQKQTKGSTALHYCCLTDNAECLKLLLRGKASIEIANESGETP 657
Db 617 QTSLSHLVDFLVQNCNLQKQTKGSTALHYCCLTDNAECLKLLLRGKASIEIANESGETP 676

Qy 658 LDIAKRLKHECEELLTQALSRFPNSHVHVEYEWLHLLHEDLDESDDDKLQPSENRRE 717

677 LDIAKRLKATQCEDLLSQAKSGKFNPHVHVEYEWLHLLHEDLDESDDDKLQPSIKKERS 736
Qy 718 DRPISFYQLGNSQLQSNVAVSLARDANLAKEKQAFMPSILQNETYCALLSG---SPPPA 774
Db 737 PRQSFCHSSSISPDQ---KLALPGFSTPRDKQRL-----SYGAFTNQIFVSTSTD 784
Qy 775 QPAAPSTTSAPPLPPRNVGK-----VQTASSANTLMKTNSVSDGGSRQRSSS 822
Db 785 SPTSP-TTEAPPLPPRNAGKGTGPTSTLPLSTQTSSTSGSSTLSKKRPPPPPPG-HKRTLS 842
Qy 823 DPPAVHPEPLPLR-----PSVMEALSQ-----PSKPAP-----VTSTNP-LTPTPPP 848
Db 843 DPRSPLPHGPNKAGVPGWNGDGGSSSKTTNKFGLSQSSSTSSAKTALGPRVLPKLPQ 902
Qy 849 PVA-KT-----PSVMEALSQ-----PSKPAP-----PGISQIRPPP 878
Db 903 KVALRKTDHLSLKDRTATIPPEIFQKSSQLAEPLQKPPPGDLPKPKTELAPKFIQIGDLPPK 962
Qy 879 -----LPPQPP-SRLPQK-----KPAQCTD 897
Db 963 GELPPKQOLGDLPPKQOLSDLPKPKQMKDLPKPKQOLGDLKAKSQTGDVSPKAAQPSVTL 1022
Qy 898 KSTP--LTNKGQPRGPVDLSATEALGFLSNAMVLQPPAPMPRKSQATKLKPKRVKALYNC 955
Db 1023 KSHPLDLSPNVQSRDAIQKQASEDSNDLTPTLP-ETVPPLPRKINTGKNKVRVKTIYDC 1081
Qy 956 VADNPDELTSSEGVIIVDGEEDQEWIGHIDGDPGRKGAFVSVFVFIAD 1006
Db 1082 QADNDELTTIEGEVIIVTGEEDQEWIGHIEGQPERKGVFPVSVFVHLS 1132

RESULT 2
US-10-104-047-2951
; Sequence 2951, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2951
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-104-047-2951

Query Match      39.3%; Score 2071.5; DB 15; Length 903;
Best Local Similarity 43.4%; Pred. No. 2.7e-108;
Matches 446; Conservative 164; Mismatches 271; Indels 147; Gaps 21;

Qy 1 MPDQISVSEFAETHEDYKAPT-ASFTTRTAQCRNTVAABEALDVRMVLKMKSVK 59
Db 1 MPEQFSVAEFLAVTAEDLSSPAGAAFAAKMPRYRGAALAREEILLEGDAILOKRIKAVR 60

Qy 60 AINSSGLAHVNEBQYTOALEKFGNCVCRDDPDLGSAFLKFSVFTKELTALFKMLIQNM 119
Db 61 AIHSSGLAHVNEBQYREAVESLGNLSHLSQNSHELSTGFLNLAVFTREVAALFKMLIQNL 120

Qy 120 NNIIISFPLDLSLKGDKLVKQKDPDKAWDYETKITKIEKKEHAKLHGMIRTEIS 179
Db 121 NNIVSFPDLSLQKLRDGRQDSKKQLEKAWDYEAAXMAKLEKE-RDRARVTGGI----- 174

Qy 180 GAETAEEMEKERRFFOLQCEYLLKVEIKIKKGVLDLQNLKYPHAQCNFFQDGLKAVE 239
Db 175 PGEVAQDMQERRIFQLHMCYELLKAGESQMKQGPDLFQSLIKFFHQAQHNFFQGWKAAQ 234

Qy 240 SLKPSIETLSTDLHTIKQADERRQLIQRLILKALQVQKESQIRQST--AYSLSHQ 297
```

```
Db 235 SUFFPIEKLAA5VHALHQAQDELOKLTQLRDSRLGTLQLESREHLSRKNSGCGYSIHQ 294
Qy 298 PQGNKEHGTERNGSLYKKSQDGIKRVWQKRCVKNQGLTISHGTANRPAPKLNLLTTCQVK 357
Db 295 HQGNKQFGTEKVGFLYKKSQDGIKRVWQKRCVKNQGLTISHGTANRPAPKLNLLTTCQVK 354
Qy 358 TNPEKKCFDLISHDRTHFOAEDBOECQIWM5VLQNSKEEALNAPFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTHFOAEDHECEAW5VLQNSKDEAL5AFGEPSAGPSGWS 414
Qy 410 TGENNIVOLTEKIISEVQRMGTNDVCCDCCGADPTWLTSTNLGILTCIETCSGHIHRELGVH 469
Db 415 AGHDEPHDLTKLLIAEVKSPGNSQCCDCCGADPTWLTSTNLGILTCIETCSGHIHRELGVH 474
Qy 470 YSPMQSLTDLVLGTSSELLAKNIGNAGFNEIMECCLPADS5VKPNP5GDMARKDYITAK 529
Db 475 FSRMQSLTDLVLGTSSELLAKNIGNAGFNEIMECCLPADS5VKPNP5GDMARKDYITAK 534
Qy 530 YIERYARKHADNAKHLSCIAEVKTRDI FGLQAYADGVDLTEKIPLANGHEPDETAL 589
Db 535 YVEHRFARR-----CTPEPQRLMTAICNRDL5VLEAFANGQDFGQPLPGDQAQPEELVL 590
Qy 590 HLA5VSDRTSLHIVDFLVQNSGNDKQTKG5TALHYCCLTDNAECLKLLRGRASIEI 649
Db 591 HLA5VQNASLPLVDFIIONGHDLDAKAADGNTALHYAALYNQPDCLKLLRGRALVGT 650
Qy 650 ANESGETPLDIARLKHCEBELTQALSGR5FN5HVHVEYEWRLHEDLDESDDDMEKL 709
Db 651 VNEAGETALDIARLKHCEBELTQALSGR5FN5HVHVEYEWRLHEDLDESDDEEEK- 709
Qy 710 QPSNRREDRPI5FYQLGSNQLQ5NA5LARDAA5LAK5KQAFMPSILQNETYGALLS- 768
Db 710 -----RCLLKLPAQAHWASGRGLDISNKTETVASL 739
Qy 769 GSPPPAQAPSTTSAPPLPRNVKQVQTASSANTLWKTNSVSV5DGG5RQRSS5DPPAVH 828
Db 740 GA-----ATPQGESEDCPPLPVKN-----SSRTLQ-----GCARHASGDRSEVS 780
Qy 829 PPLPLRVSTNPLTPTPPPVAKTP5VMEALSQPSKP--APPGISQ--IRPPLPPQP-- 883
Db 781 -----SLSSEAPETPESLGS5P5SS5L5M5PL--EPGDP5QAPPN5EGLR5EPGTSR5PL 833
Qy 884 -----PSRLPQKKPAPGTDKSTPLTNKQ5PGR5VDLSATEALG5PL5NAMVLO5PPAP5PRK 938
Db 834 TSGTTPSEM--YLPVRF5SESTR5YRG-ARSPED-----GPSAR-----QPLPRR 876
Qy 939 SQATKLPKRVKALNCVADNPDELTFSEGDV5I5VDGEDQEW5IGHIDG5PGRK5APV 998
Db 877 NVP-----VGITEGDSGRTGSLPA 895
Qy 999 SPVHFIAD 1006
Db 896 SSVQLIQD 903
```

RESULT 3

```
US-10-490-605-2
; Sequence 2, Application US/10490605
; Publication No. US20050019768A1
; GENERAL INFORMATION:
; APPLICANT: JAPAN AS REPRESENTED BY THE PRESIDENT OF THE UNIVERSITY OF TOKYO
; APPLICANT: ONCOTHERAPY SCIENCE, INC.
; TITLE OF INVENTION: HEPATOCELLULAR CARCINOMA-RELATED GENES AND POLYPEPTIDES, AND METH
; TITLE OF INVENTION: FOR DETECTING HEPATOCELLULAR CARCINOMAS
; FILE REFERENCE: 25371-029NATL/SEN-A0121P-US
; CURRENT APPLICATION NUMBER: US/10/490,605
; PRIOR FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: USSN 60/324,261
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: CA
; PRIOR FILING DATE: 2002-08-23
; NUMBER OF SEQ ID NOS: 47
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-490-605-2

Query Match 39.3%; Score 2071.5; DB 17; Length 903;
Best Local Similarity 43.4%; Pred. No. 2.7e-108;
Matches 446; Conservative 164; Mismatches 271; Indels 147; Gaps 21;
```

```
Qy 1 MPDQISVSEFAVETHEDEYKAPT-ASSPTTRTAQACNTVAATEALD5VDRVLYKMK5VK 59
Db 1 MPEQSVSAEFLAVTAEDLSSPAGAAFAAKMPRYRGAALAREEILLEGQAIILQRIKAVR 60
Qy 60 AINSGLAHVNEEQYTQALBKFGCNCVCRDDPDLGSAFLKPSVFTKELTALFKNLQNM 119
Db 61 AIHSSGLGHVNEEQYREAVESLGN5HLSQNSHSLSTGFLNLA5VTRVLAALFKNLQNL 120
Qy 120 NNI5PPLD5LLKGLDKGVKGLDKKPPDKAWKDYETKITKEKEK5HAKLHGMTRTEIS 179
Db 121 NNIV5PPLD5LLKGLDKGVKGLDKKPPDKAWKDYETKITKEKEK5HAKLHGMTRTEIS 174
Qy 180 GAETAEMEKEKRRFPOLQMCVLLK5VNEIKK5GV5DL5LQNLK5YF5AQC5NFFQDGLKAVE 239
Db 175 PGEVAQDMORERRIFQLHMCBYLLKAG5SOMKQ5PDFLQ5LLK5F5HAQ5NFFQDGLKAAQ 234
Qy 240 SLKPSIETSLDHLTIKAQDEERRQLTLQRLDILK5ALQVEQK5QIRQST--AYSLHQ 297
Db 235 SLFPIEKLAA5VHALHQAQDELOKLTQLRDSRLGTLQLESREHLSRKNSGCGYSIHQ 294
Qy 298 PQGNKEHGTERNGSLYKKSQDGIKRVWQKRCVKNQGLTISHGTANRPAPKLNLLTTCQVK 357
Db 295 HQGNKQFGTEKVGFLYKKSQDGIKRVWQKRCVKNQGLTISHGTANRPAPKLNLLTTCQVK 354
Qy 358 TNPEKKCFDLISHDRTHFOAEDBOECQIWM5VLQNSKEEALNAPFKGD-----DN 409
Db 355 PNPEKKCFDLVTHNRTHFOAEDHECEAW5VLQNSKDEAL5AFGEPSAGPSGWS 414
Qy 410 TGENNIVOLTEKIISEVQRMGTNDVCCDCCGADPTWLTSTNLGILTCIETCSGHIHRELGVH 469
Db 415 AGHDEPHDLTKLLIAEVKSPGNSQCCDCCGADPTWLTSTNLGILTCIETCSGHIHRELGVH 474
Qy 470 YSPMQSLTDLVLGTSSELLAKNIGNAGFNEIMECCLPADS5VKPNP5GDMARKDYITAK 529
Db 475 FSRMQSLTDLVLGTSSELLAKNIGNAGFNEIMECCLPADS5VKPNP5GDMARKDYITAK 534
Qy 530 YIERYARKHADNAKHLSCIAEVKTRDI FGLQAYADGVDLTEKIPLANGHEPDETAL 589
Db 535 YVEHRFARR-----CTPEPQRLMTAICNRDL5VLEAFANGQDFGQPLPGDQAQPEELVL 590
Qy 590 HLA5VSDRTSLHIVDFLVQNSGNDKQTKG5TALHYCCLTDNAECLKLLRGRASIEI 649
Db 591 HLA5VQNASLPLVDFIIONGHDLDAKAADGNTALHYAALYNQPDCLKLLRGRALVGT 650
Qy 650 ANESGETPLDIARLKHCEBELTQALSGR5FN5HVHVEYEWRLHEDLDESDDDMEKL 709
Db 651 VNEAGETALDIARLKHCEBELTQALSGR5FN5HVHVEYEWRLHEDLDESDDEEEK- 709
Qy 710 QPSNRREDRPI5FYQLGSNQLQ5NA5LARDAA5LAK5KQAFMPSILQNETYGALLS- 768
Db 710 -----RCLLKLPAQAHWASGRGLDISNKTETVASL 739
Qy 769 GSPPPAQAPSTTSAPPLPRNVKQVQTASSANTLWKTNSVSV5DGG5RQRSS5DPPAVH 828
Db 740 GA-----ATPQGESEDCPPLPVKN-----SSRTLQ-----GCARHASGDRSEVS 780
Qy 829 PPLPLRVSTNPLTPTPPPVAKTP5VMEALSQPSKP--APPGISQ--IRPPLPPQP-- 883
Db 781 -----SLSSEAPETPESLGS5P5SS5L5M5PL--EPGDP5QAPPN5EGLR5EPGTSR5PL 833
Qy 884 -----PSRLPQKKPAPGTDKSTPLTNKQ5PGR5VDLSATEALG5PL5NAMVLO5PPAP5PRK 938
Db 834 TSGTTPSEM--YLPVRF5SESTR5YRG-ARSPED-----GPSAR-----QPLPRR 876
Qy 939 SQATKLPKRVKALNCVADNPDELTFSEGDV5I5VDGEDQEW5IGHIDG5PGRK5APV 998
Db 877 NVP-----VGITEGDSGRTGSLPA 895
Qy 999 SPVHFIAD 1006
Db 896 SSVQLIQD 903
```



```
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2284

Query Match      20.4%; Score 1075.5; DB 15; Length 208;
Best Local Similarity 81.4%; Pred. No. 6.5e-53;
Matches 206; Conservative 1; Mismatches 1; Indels 45; Gaps 1;

Qy 754 MPSILQNETYGALLSGSPPPAQAAPSTTSAPPLPRNVGVKQTASSANTLWKTNSVSD 813
Db 1 MPSILQNETYGALLSGSPPPAQAAPSTTSAPPLPRNVGK----- 41

Qy 814 GGSQRSSDPPAVHPPPLPLRVTSNPLTTPPPPVAKTSPVMEALSQPSKPAPPGISQ 873
Db 42 -----DPLTTPPPPVAKTSPVMEALSQPSKPAPPGISQ 75

Qy 874 IRPPLPQPSRLPQKYPAGTDKSTPLTNKGQPRGVDLSATEALGPLSNAMVLQPPA 933
Db 76 IRPPLPQPSRLPQKYPAGTDKSTPLTNKGQPRGVDLSATEALGPLSNAMVLQPPA 135

Qy 934 PMPRKSQATKLKPKRKVKALNCVADNPDELTFSEGDVLIIVDGEEDQEWIGHIDGDPGRK 993
Db 136 PMPRKSQATKLKPKRKVKALNCVADNPDELTFSEGDVLIIVDGEEDQEWIGHIDGDPGRK 195

Qy 994 GAPPVSVFHFAD 1006
Db 196 GAPPVSVFHFAD 208

RESULT 7
US-10-408-765A-2815
; Sequence 2815, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Sojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2815
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2815

Query Match      11.6%; Score 611.5; DB 16; Length 358;
Best Local Similarity 43.0%; Pred. No. 1.9e-26;
Matches 128; Conservative 17; Mismatches 18; Indels 135; Gaps 1;

Qy 68 HVNEEQVTALEKFGNVCRRDDPLGSAFLKFSVFTKELTALFKNLI----- 116
Db 5 HVQNEENAVQLDKFGSNFLSRDNDLGTAFVKPSTLTKELSTLKNLVRHFIFVRNAV 64

Qy 117 ----- 116
Db 65 TSAEQSNYGRKMIEDDPSTCKIFCLLHPRSLYPKATTPDLDCWEVLQPSLVLFSGIM 124
```

```
Qy 117 ----- 116
Db 125 LASTGKHGLVRGAWLDFRLPLYKPGVLEVPFVLCSEALLSEARKNEALLTDCLSASWLLQ 184

Qy 117 ---QNNNNIISFPLDSLLKGLAGVKGDLLKPPDKAWKQVETKITKIEKKEKHAHLHG 172
Db 185 QDRLOGLSHNVIFTLDSLLKGLAGVKGDLLKPPDKAWKQVETKITKIEKKEKHAHLHG 244

Qy 173 MIRTEISGAETAEEMEKERRFFQLQMCCEYLLKVNNEIKIKGVDLLQNLIKYFHAQCNF 230
Db 245 MIRTEITGAETAEEMEKERRLFLQMCCEYLLKVNNEIKIKGVDLLQNLIKYFHAQCNW 302

RESULT 8
US-10-080-334-238
; Sequence 238, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
```

```
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 238
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-238
```

```
Query Match 10.8%; Score 569.5; DB 15; Length 778;
Best Local Similarity 24.6%; Pred. No. 1.2e-21;
Matches 197; Conservative 138; Mismatches 301; Indels 165; Gaps 26;

Qy 38 AAIEAL-DVDRMVLVYKMKSKVAINSSGLAHVENEEOYTOALEKFGNGCVRD----- 90
Db 19 AALEVEGDVALEL-KLDKLVKL-----GIAMIDTGKAFVANKQFMNG--IRDLAQYSS 71
Qy 91 -DPDLGSAFLKFSVFTKELTAFKNLIQNMNIIISFPLDSLKGDLKGVKGLKPKPDKA 149
Db 72 NDAVETSLTKFSDLSQEMINFTILLPNSSEINLRHSFNFVKEDLRKPK-DAAKQFEKV 130
Qy 150 WKDYETKITK---TEKEKEHAKLHGMTRTISGAIEAEMEKEKRRFFQLOMCEYLLKVN 206
Db 131 SEKENALVKNQAQVQRNQHVE-----EATNLTATKCFRHALDYVLQIN 178
Qy 207 EIKIKGVLDLQNLKIKYFAOCNFFQDGLKAVESLKPSIETLTDLHTIKQAQDERROL 266
Db 179 VLQSKRSEILKSMISFYAHIAFHQGYDIFSELGPTWKDLGAQLDRLVGDAAKEKREM 238
Qy 267 IQLRDILKALQVEQKDSQIRQSTAYSILHQPQGNKEHGTNRNGLYKKSQGIKRVKQKR 326
Db 239 EQKHS-----TIQKDFSRDSDSLKLVN-----DAANGIVMEGYLFKRASNAFKTNRR 287
Qy 327 KCSVKNGFLTSHGTANRPPAKL-NLLTCQVK--TNPEKKCFDLISHDRYTHFOAEDEQ 383
Db 288 WFSIQNNQVYQKKFKDNPVTVVEDRLCTVKHCEDIERRFCEVVSPTKSCMLQADSEK 347
Qy 384 EQIIMSVLQNSKEBALNNAFKGDD-----NTGE-----NNIVOBELTK--EIISEV 427
Db 348 LRQAWIKAVQTSIATAYRE--KGDSEKLDKSSPSTGSLGNSKEKLLKGESALQRV 405
Qy 428 QRMGTNDVCCDGPDPPTWLSNLGILTCIECGIHRBLGVHYSPMSQLTLDVLGTSBLL 487
Db 406 QCIPGNASCCDGLADPRWASINLGTILCIECGIHRSLGVHFSKVRSLTLDTWPELLK 465
Qy 488 LAKNIGNAGFNEIMECCLPADSVPKPNFGSDMNARKDVIYAKYIERRYA----- 536
Db 466 LMCELGNDVINRVYEANVEKMGIKKPPQG-QRQEKEAYIRAKYVERKFDVKYISLSLSPPE 524
Qy 537 -RKKHADNAAKLHSLC-----EAVKTRDIFGLQAYADGVLDLITEKIPLAN 580
Db 525 QOKKFKVSKSSEKRLSISKFGPGDQVRASQAQSVRSND-SGIQQSSDDGRESLPSTVSAN 583
Qy 581 G-HEPD-----ETALHLA----- 592
Db 584 SLYPEGEGRQDSSMFLDSKHLNPGQLYRASYEKNLPKMAEALAHGADVNNANSENKAT 643
Qy 593 ---RVSDRTSLHIVDFLVQNSNLQKQTKGKSTALHYCCLTDNASCLKLLKGKASIEIA 650
Db 644 PLIQAVLGSLVTCBFLQNGANVQRDVQGRGPLLHATVLGHTQGVCLFLKRGANQHAT 703
Qy 651 NESGTPDLIDAKRLKHECEELLTQALSGRFNSHVHVEYEWELLHEDLDESDDMDKEL- 709
Db 704 DEEGKPLSIAVEANADIVTLRLA-----RMNEEMRES-----EGLY 742
Qy 710 -QPSNRRREDRPISTFYQLGSN 729
Db 743 QQPGDETQDIPRDFSQMASN 763
```

```
RESULT 9
US-10-080-334-80
; Sequence 80, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Svytek, Kimberly A
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zethusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-80
```



```
Query Match 10.6%; Score 558; DB 15; Length 764;
Best Local Similarity 23.2%; Pred. No. 5.2e-23;
Matches 174; Conservative 119; Mismatches 315; Indels 142; Gaps 15;

QY 26 FTTRTAQCRNTVAALBEALDVRMVLVYKMKSVKAINSSGLAHVENEEOYTQALEKFGN 85
Db 15 FRATIDEVETDVEIEAKLD-----KLVKLCSGMVEAGKAYVSTSRFLVSGVDRLSQQ 67

QY 86 CVCRRDDPLGSAFLKPSVETKELTALFKNLIQNMNIIISFPDLSLLKGLKGVKGLKRP 145
Db 68 --COGDTVISECLQRFADSLQEVVNMILFDQAQSVRQQLQSFVKEDVRKFK-ETKKQ 124

QY 146 FDKAKVDYETKTKIEKEKEHAKLHGIMRTISGAETAEENEKERRFPOLQMCYLLKV 205
Db 125 FDKVREDLELSLV-----RNAQAPRRHPHEVE-----EATGALTITRKCFRHLALDYVLQI 175

QY 206 NEIKTKGVLDLQNLKIFPHACNFFQDGLKAVESIKPSIETLSLDTLHTIKQAQDEERQ 265
Db 176 NVLOAKKFEILDMSLFWHAQSSFPQGYSLHQDPYMKLAELDQLVDSAVEKRE 235

QY 266 LIQLRDILKSALQVQKEDSIQIRQSTAYSLHQPOGNKE-----HGTERNGSLYKSDGI 319
Db 236 M-----ERKHAALQORTLRDPSYDESKVEFDVDPAGSVVMVEGLFKRASNA 281

QY 320 RKWOKRCKSVKQGLFTTISHGTANRPPAKL-NLLTCQVK--TNPEKKCFDLISHDRYH 376
Db 282 FKTNRWRFSIQNSLVYQKKLKDALITVVVDLRLCSVKPCDIERRFCFEVLSPKSCM 341

QY 377 FOAEDQEQCIWMSVLQNSKEAL-----NNAPKDDNTGNNIVQ 417
Db 342 LQADSEKURQAWQAVQASIASAYRESPOSCYSERLDRTASPTSSIDSATDTRERGKVG 401

QY 418 ELTKIIEVQRTGNDVCCDCAPTWLSNLGILTCIEGSHRELGHVHSPQSLT 477
Db 402 E--SVLQVQSVAGNSQCGDQCPDPRWASINLGVLLCIEGSHRSLGVHCSKVRSLT 458

QY 478 DLVLGTSELLAKNTIGNAFNIMECCUPAEDSVKPNPQSGDMNARKYITAKYIERRYAR 537
Db 459 LDSWEPELLKLMCELGSAVNQIYEAQCEGAGSRKPTASSSRQDKEAMIKDYVEKKFLR 518

QY 538 K-----AVKTRDIFGLQAYADGVDLTEKIPLANGHEPDE 586
Db 539 KAPMAPALEAPRRVRVQKLRPHSSPRAPTARKVRLEPVLPCVAALSSEGAESSESGE 578

QY 539 KHADNAAKLHSLCE-----AVKTRDIFGLQAYADGVDLTEKIPLANGHEPDE 586
Db 579 ADGDTAEAWGLADVREHLPGLLAHRAARADLPALAAALAHGAE-----VNWADAEDSGK 634

QY 587 TALHLAVRSVDRTSLSHIVDFVQNSGNLDKQTKGKSTALHYCCLTDNAECLKLLRGKAS 646
Db 635 TPL---VQAVLGSLIVCFELFLQNGADVNRQDSRGRAPLHATLLGRTGQVCLFKRGAD 691

QY 647 IEIANESGETPLDIKRLKHECEBLLTQA 676
Db 692 CHALDQEQRDPLAIAVQAANADIVTLRLIA 721
```

```
RESULT 10
US-10-080-334-235
; Sequence 235, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkete, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J

Query Match 10.4%; Score 551; DB 15; Length 804;
Best Local Similarity 22.5%; Pred. No. 1.4e-22;
Matches 180; Conservative 119; Mismatches 323; Indels 178; Gaps 16;

QY 12 AETHEDYKAPTASSFTTRTAQCRNTVAALBEALDVRMVLVYKMKSVKAINSSGLAHVEN 71
Db 5 AEAAGDSQP-RSCFRATIDEVETDVEIEAKLD-----KLVKLCSGMVEAGKAYVST 56

QY 72 BEQYTOALEKFGNVCVCRDDPDLSGSAFLKPSVFTTKELTALFKNLIQNMNIIISFPDLSLL 131
Db 57 SRLFVSGVDRLSQQ--COGDTVISECLQRFADSLQEVVNMILFDQAQSVRQQLQSFV 114

QY 132 KGDLLGVKGLKPKFPDKAWKDYETKTKIEKEKEHAKLHGIMRTISGAETAEENEKERR 191
Db 115 KEDVRKFK-ETKKQFDKVRDELELSLV-----RNAQAPRRHPHEVE-----EATGALTITR 164

QY 192 RFFQLQMCYEYLLKYNEIKKGVLDLQNLKIFPHACNFFQDGLKAVESLKPSTETLSTD 251
```


; Sequence 237, Application US/10080334		
; Publication No. US20040002584A1		
; GENERAL INFORMATION:		
; APPLICANT: Pena, Carol E. A.		
; APPLICANT: Shimkets, Richard A		
; APPLICANT: Li, Li		
; APPLICANT: Shenoy, Ramesh G		
; APPLICANT: Kekuda, Ramesh		
; APPLICANT: Spytek, Kimberly A.		
; APPLICANT: Vernet, Corine A. M.		
; APPLICANT: Malyankar, Uriel M		
; APPLICANT: Guo, Xiaojia		
; APPLICANT: Gusev, Vladimir Y		
; APPLICANT: Casman, Stacie J		
; APPLICANT: Boldog, Ferenc L		
; APPLICANT: Furtak, Katarzyna		
; APPLICANT: Tchernev, Velizar T		
; APPLICANT: Patturajan, Meera		
; APPLICANT: Gangolli, Esha A		
; APPLICANT: Padigaru, Muralidhara		
; APPLICANT: Liu, Xiaohong		
; APPLICANT: Baumgartner, Jason C.		
; APPLICANT: Gerlach, Valerie		
; APPLICANT: Spaderna, Steven K		
; APPLICANT: Zerhusen, Bryan D		
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of		
; FILE REFERENCE: 21402-275		
; CURRENT APPLICATION NUMBER: US/10/080,334		
; PRIOR FILING DATE: 2002-02-21		
; PRIOR APPLICATION NUMBER: 60/270,523		
; PRIOR FILING DATE: 2001-02-21		
; PRIOR APPLICATION NUMBER: 60/322,712		
; PRIOR FILING DATE: 2001-09-17		
; PRIOR APPLICATION NUMBER: 60/311,980		
; PRIOR FILING DATE: 2001-08-13		
; PRIOR APPLICATION NUMBER: 60/330,307		
; PRIOR FILING DATE: 2001-10-18		
; PRIOR APPLICATION NUMBER: 60/278,796		
; PRIOR FILING DATE: 2001-03-26		
; PRIOR APPLICATION NUMBER: 60/281,521		
; PRIOR FILING DATE: 2001-04-04		
; PRIOR APPLICATION NUMBER: 60/276,677		
; PRIOR FILING DATE: 2001-03-16		
; PRIOR APPLICATION NUMBER: 60/311,595		
; PRIOR FILING DATE: 2001-08-10		
; PRIOR APPLICATION NUMBER: 60/270,220		
; PRIOR FILING DATE: 2001-02-21		
; PRIOR APPLICATION NUMBER: 60/274,295		
; PRIOR FILING DATE: 2001-03-08		
; PRIOR APPLICATION NUMBER: 60/318,526		
; PRIOR FILING DATE: 2001-09-10		
; PRIOR APPLICATION NUMBER: 60/286,548		
; PRIOR FILING DATE: 2001-04-25		
; PRIOR APPLICATION NUMBER: 60/291,765		
; PRIOR FILING DATE: 2001-05-17		
; PRIOR APPLICATION NUMBER: 60/270,797		
; PRIOR FILING DATE: 2001-02-23		
; PRIOR APPLICATION NUMBER: 60/276,400		
; PRIOR FILING DATE: 2001-03-16		
; PRIOR APPLICATION NUMBER: 60/270,810		
; PRIOR FILING DATE: 2001-02-23		
; NUMBER OF SEQ ID NOS: 388		
; SOFTWARE: PatentIn Ver. 2.1		
; SEQ ID NO 237		
; LENGTH: 834		
; TYPE: PRT		
; ORGANISM: Homo sapiens		
US-10-080-334-237		
Query Match		10.0%; Score 529; DB 15; Length 834;
Best Local Similarity		21.4%; Pred. No. 2.5e-21;
Matches 175; Conservative 119; Mismatches 315; Indels 210; Gaps 15;		
Qy	26	FTTRTAQCRNTVAALBEALDVRMLVKMKSVKAINSSGLAHVNEEQYTQALEKEGCGN 85
Db	15	PRATIDEVETDVVEIEAKLD-----KLVKCSGWEAGKAYVSTSRVSGVGRDLSQQ 67
Qy	86	CVCRRDDPLGSAFLKFSVFTKELTALPKNLQNNNIISFPLDLSLLGDKLKGVDLKKP 145
Db	68	--CQGDVISECLQRFPADSLQEVVNYHMILFDQAQSRVROQLQSFVKEDVRKFK-ETKKQ 124
Qy	146	PDKAWKDYETKIKIEKEKEHAKLHGMIRTEISGALIAEEMEXERFFQLOMCEYLLKV 205
Db	125	FDKVRDELLESLV-----RNAQAPRHPHEVE-----EATGALTLTRKCFRHLALDYVLQI 175
Qy	206	NEIKIKKGVLLQNLIKYFHAQCNFFQDGLKAVESLKPSITLTDL-----HTIKQAQ 259
Db	176	NVLOAKKKFEILDLSNLSFMRAQSSFFQGGYSLHLQLDPYMKKAELDLQVIDSAVEKRE 235
Qy	260	DEERRQLIQRLDILKSALQVEKEDSQIROSTAYSLHQPOGNKEHGTGERSLSYKKS DGI 319
Db	236	MERKHAALQQTLLQDFSYDESKVEFDV-----DAPSGVVMGSLFKRASNA 282
Qy	320	RKVMQKRCVKNGFLTISHGTANRPPAKL-NLLTCQVK--TNPEKKKCPDLISHDRYH 376
Db	283	FKTWNRWFSTQNSQLVYQKKLDALTVMVDDLRLCSVKPCEDIERRFCPEVLSPTKSCM 342
Qy	377	FOADEQECCQIMSVLONSKEAL-----NNAFKGDNDTGNENIVQ 417
Db	343	LQADSEKLRQAWQAVQASIASAYRESPDCYSERLDRTPASTSSIDSATDTRGVKG 402
Qy	418	ELTKEIIEVQRTMGNDVCCDGPDTWLSTNLGILTCIECSGIRHRELGVHYSPMOSLT 477
Db	403	E---SVLQRVQSVAGNSQCGDQGPDRWASINLGVLCTIECSGIRHSLGVHCKVRSLT 459
Qy	478	LDVLGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYTIERRYAR 537
Db	460	LDSWEPELLKMLCGLNSAVNQIYEAQCEGAGSRKPTASSRQDKBAWIKDKYVEKKPLR 519
Qy	538	K-----
Db	520	KAPMAPALEAPRRWRVQKCLRPHSSPRAPTARRKVRLEPVLPCVAALSSVGLDRKFRD 579
Qy	539	-----
Db	580	SLFCPELDLSLYFDAGAGAGAPRSLSSDGLGSSDGLAFSGSVVDSVTEBEG 639
Qy	539	-----KHADNAKHLSLCE-----AVKTRDIFGLLQAYAGVDLTETKIP 577
Db	640	AESESSGEADGDTAEAWGLADYRELHPGLLAHRAARADLPALAAALAHGAE-----VN 695
Qy	578	LANGHEPDETALHLAVRSVDRSTSLHIVDFLVQNSGNDLKOTGKGTALHYCCLTDNABCL 637
Db	696	WADADEGKTPL---VQAVLGSLIVCEFLQNGADVNRQDSRGRAPLHATLLGRIGQV 752
Qy	638	KLLLRGKASIBIANESGETPLDIARLKHCEHELLTQA 676
Db	753	CLFLKRGADQHALDQEQRDPLAIAVQAANADIVTLRLA 791
RESULT 14		
US-10-467-434-4		
; Sequence 4, Application US/10467434		
; Publication No. US20040092715A1		
; GENERAL INFORMATION:		
; APPLICANT: DING, Li; WARREN, Bridget A.;		
; APPLICANT: ELLIOTT, Vicki S.; TANG, Y. Tom;		
; APPLICANT: YUE, Henry; BURFORD, Neil;		
; APPLICANT: LEE, Sally; RICHARDSON, Thomas W.;		
; APPLICANT: LAL, Preeti G.; NGUYEN, Dannel B.;		
; APPLICANT: YANG, Junning; HAFALIA, April J.A.;		
; APPLICANT: YONG, Craig H.; GURURAJAN, Rajagopal;		
; APPLICANT: BAUGHN, Mariah R.; WANG, Yumei E.;		
; APPLICANT: YAO, Monique G.; THANGAVELU, Kavitha;		
; APPLICANT: SWARNAKAR, Anita; GRIFFIN, Jennifer A.;		

APPLICANT: FORSYTHE, Ian J.; EMERLING, Brooke M.;
APPLICANT: CHAWLA, Narinder K.;
TITLE OF INVENTION: Intracellular Signaling Molecules
FILE REFERENCE: PF-0897 USN
CURRENT APPLICATION NUMBER: US/10/467,434
PRIOR FILING DATE: 2003-08-06
PRIOR APPLICATION NUMBER: PCT/US02/03966
PRIOR FILING DATE: 2002-02-07
PRIOR APPLICATION NUMBER: US 60/267,925
PRIOR FILING DATE: 2001-02-08
PRIOR APPLICATION NUMBER: US 60/274,435
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: US 60/277,819
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/281,326
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: US 60/291,195
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/291,550
PRIOR FILING DATE: 2001-05-16
PRIOR APPLICATION NUMBER: US 60/293,591
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/295,348
PRIOR FILING DATE: 2001-06-01
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PERL Program
SEQ ID NO 4
LENGTH: 834
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 6892116CD1
US-10-467-434-4

Query Match 10.0%; Score 529; DB 15; Length 834;
Best Local Similarity 21.4%; Pred. No. 2.5e-21;
Matches 175; Conservative 119; Mismatches 315; Indels 210; Gaps 15;

QY 26 FTTRTAQCRNTVAATEEALDVRMVLVYKMKSVKAINSSGLAHVNEBQYTOALEKFGCN 85
DB 15 FRATIDEVETDVEAEKLD-----KLVKLCGMEAGKAYVSTSLFVSGVNDLSQQ 67
QY 86 CVCDDPDLGSAFLKFSVFTKELTALFKNLQNNNNIISFPLDSLLKGLDKGVKGLKKP 145
DB 68 --COGDTVISECLORFADSLQEVVNYHMLFDQAQSRVQQLQSPVKEDVRKEK-ETKQ 124
QY 146 FDKAWDYETKITKEKKEHAKLHGMRITISGAETAEEMEKERRFFQLOMCEYLKV 205
DB 125 FDKVREDLESLV-----RNAQAPRRPHEVE---EATGALTTRKCFRHLALDYVLQI 175
QY 206 NEIKIKGVLDLONLKYFHAQCNFFQDGLKAVESLKSITLSTDL-----HTIKQAQ 259
DB 176 NVLQAKKFEILDMSLWFMHQAQSFQOYSLHQLDPMYKLAELDLQVLIDSAVEKRE 235
QY 260 DEERQLQLRDLILKALQVEKEDSIQSTAYSLHQPQNGKHGTERNGSLYKKS DGI 319
DB 236 MERKHAALQORTLLQDPSYDESKVEFDV-----DAPSGVMEGYLFKASNA 282
QY 320 RYVQKRCVKNQGLTTSHTANRPPAKL-NLITCQVK--TNPEEKCFDLISHDRYH 376
DB 283 FKTWNRWFSIQNSOLVYQKLLKDALTVVDDLRCLSVKPCDIERRFFCVLSPTKSCM 342
QY 377 FOADEQECQIWMVLONSKEAL-----NNAPKGDNDTENNIVQ 417
DB 343 LOADESEKLQAWQVQVQASIASAVNESPDSCYSEKRLDTASPTSSIDATDTRERGK 402
QY 418 ELTKELIIEVQRMGTNDVCCGADPDPTWLTSLNLTGILTCIECSGTHRELGVHYSPMSLT 477
DB 403 E---SVLQVQSVAGNSQCGCGQDPRWASINLGVILCIECSGTHRELGVHYSPMSLT 459
QY 478 LDVLGTSSELLAKQIGNAGFNEIMECCPLPAEDSVKPNPFGSDMNAKQDITAKYIERRYAR 537

DB 460 LDSWPEPELLKLMCELGNSAVNQIYEAQCEGAGSRKPTASSSRQDKBEAWIKQYVEKKFLR 519
QY 538 K-----
DB 520 KAPWAPALEAPRRWRVQKCLRPHSSPRAPTARRKVRLEPVLPCVAALSSVGTLDKRFRRD 579
QY 539 -----
DB 580 SLFCFDELDSLFSYFDAGAAGAPRSLSSDSGLGSSDGLAFSGSVVDSVTEBEG 639
QY 539 -----KHADNAAKLHSLCE-----AVKTRDIFGLQAYADGVLDLTKIP 577
DB 640 ABESSESGEADGTEAEAWGLADVRELHPGLLAHRAARADLPALAAALAHGAE-----VN 695
QY 578 LANGHEPDETALHLAVRSVDRSTSLHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECL 637
DB 696 WADADEGKTP-----VQAVLGSLIVCEFLQNGADVNRDSRGRAPLHATLLGRTQV 752
QY 638 KLLLRGKASIEIANESGETPLDIAKRLKHEHCEBELLTOA 676
DB 753 CLFLKRGADQHALDQEQRDPLAIAVQAANADIVTLRLA 791

RESULT 15

US-10-739-930-6184
Sequence 6184, Application US/10739930
Publication No. US20040216190A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
FILE REFERENCE: 38-21(53377)B
CURRENT APPLICATION NUMBER: US/10/739,930
CURRENT FILING DATE: 2003-12-18
NUMBER OF SEQ ID NOS: 11088
SEQ ID NO 6184
LENGTH: 775
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: Clone ID: APATH-23APR03-C24093_1.p
US-10-739-930-6184

Query Match 9.7%; Score 509.5; DB 16; Length 775;
Best Local Similarity 24.3%; Pred. No. 2.9e-20;
Matches 190; Conservative 123; Mismatches 292; Indels 177; Gaps 24;

QY 37 VAAITEALDVRMVLVYKMKSVKAINSSGLAHVNEBQYTOALEKFGNCVCRDDPDL-- 94
DB 17 VCSLEGTADLKDRCQKLYKGVKFMGTGLGASKGESAFACLEBFGGG---PDDPISLS 73
QY 95 --GSAFLKFSVFTKELTALFKNLQNNNNIISFPLDSLLKGLDKGVKGLKKPDKAWD 152
DB 74 IGGPVISKFINAURELASYKFEFLCSQVHVLLELMNPFISVDLOEAK-ESRHRFDKAAHS 132
QY 153 YETKITKEKKEHAKLHGMRITISGAETAEEMEKERRFFQLOMCEYLKLVNEIKK 212
DB 133 YDQSRKFSVLSKN-----TRGEIV-ABLEEDLENSKSTFEKSRNLVNSLTIEAK 184
QY 213 GVDLLQNLKYFHAQCNFFQDGLKAVESLKSITLSTDLHTIKO-----AQDEERRQLIQ 268
DB 185 KYEFLSISAIMDAHLRYFKLGYDLLQLEPFIHQILTYAQQSKESQKIEQDLRLARIQE 244
QY 269 LRDILKALQVEKEDSIQSTAYSLHQPQNGKHG-----TE-----RNGSL 312
DB 245 FR-----TQSELDSSQLVANAES--SGANGNRVGNIPYKNTETSLTADKEVIKQYL 295
QY 313 YKXSDGIRKVMQKCKSVKNGFLTISHGTA-----FVLDHSGMYVYRTNKGMSHHHSGSSDHTGVGR 342
DB 296 LKRSSSURTW-KKKF-----FVLDHSGMYVYRTNKGMSHHHSGSSDHTGVGR 349
QY 343 ----NRPPA-----KLNLLTCOVKTNPB---KCCFDLISHDRTYHFOADEBOEC 385

Db 350 FRARHNRSGSLTEGSLGYNTIDLRSLIKLDAEDMDLRLCFRIISPOKTYTLQAENGADR 409
Qy 386 QIWMSVLQNSKEEALN-----NAPKGDNTGENNIVQELTKEI 423
Db 410 MDWVNKITKAIGTLNLSHFLOQSPVRYLDKONSSAPANAVVSGQIRHNSRQNIIGDDV 469
Qy 424 ISEVORTGNBVCDCGAPDPTWLSNLTGILTCIECSGIHRELGVHYSQMOSLTLD--VL 481
Db 470 STILRGLFGNNAECNAPEDPWASLNLGVLLCQCSCGVHRNLGVHISKVRSLSLDVKW 529
Qy 482 GTSELLLLAKNIGNAGFNEIMECCLPAEDSV-----KPNPGSDMMARKDYITAK 529
Db 530 EPTILDLPRLNGVYCNLSLWEGLLHLDLDDCEDGSALSASVSKPCPEDSFSVKKYILGK 589
Qy 530 YIERRYARKHAD-NAAKHLSLCEAVKT---RDI FGLLOAYAD----- 568
Db 590 YLEKALVIKDESEANLSAASRIWEAVQSRNIREIYRLIVTTGVDVNIINTKFDIDTIDAY 649
Qy 569 -GVDLTEKIPLANGHEPDE-----TALHLAVRSVDRTSLSLHIVDFLVQN 610
Db 650 HHIDAEEK-AVKKRHDPTVCORIKESNEPRSCLOQCSLLHVACHIGDSV---LLELLLOP 705
Qy 611 SGNLDKQTKGSTALHYCLTDNAECLKLLLRGKASIEIANESGETPLDIAKRLKHECE 670
Db 706 GADLNIRDYHGRTPLHHCISSGNHKFAKILLRRGARPSTIEDDGGLSVLERAMEMGAI 765
Qy 671 EL 672
Db 766 EL 767

Search completed: August 4, 2005, 08:34:44
Job time : 108 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 16:31:46 ; Search time 7497 Seconds
(without alignments)
5107.730 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAPFVSVHFHFIAD 1006

Scoring table:

OLIGO Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Word size: 1

Total number of hits satisfying chosen parameters: 68477535

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US0914042/runat_04082005_090747_2671/app_query.fasta_1.1159
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0914042 @CGN 1 1 4326 @runat_04082005_090747_2671 -NCPU=6 -ICPU=3
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsel1:*
9: gb_gsel2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	202	20.1	1167	5	BU508193
2	198	19.7	901	5	BU172807
3	177	17.6	556	1	AL044307
4	165	16.4	846	5	BU155022
5	160	15.9	689	5	BU703510
6	160	15.9	731	7	CN538273
7	160	15.9	773	7	CN460843
8	160	15.9	781	7	CO429713
9	157	15.6	582	5	BP280849

10	156	15.5	485	4	BM839307	BM839307	K-EST0116
11	154	15.3	702	7	CF532249	CF532249	UI-M-GHO-
12	154	15.3	884	5	BQ439046	BQ439046	AGENCOURT
13	151	15.0	948	6	CD358314	CD358314	AGENCOURT
14	148	14.7	714	7	CF745192	CF745192	UI-M-GVO-
15	148	14.7	742	7	CN535077	CN535077	UI-M-HSO-
16	148	14.7	1014	4	BG293568	BG293568	602390218
17	143	14.2	738	7	CN537253	CN537253	UI-M-HSO-
18	142	14.1	737	7	CF539222	CF539222	UI-M-GHO-
19	141	14.0	702	6	CD350751	CD350751	UI-M-GIO-
20	137	13.6	514	5	AW996731	AW996731	QV3-BN004
21	133	13.2	483	5	BX476938	BX476938	DKF2686C
22	132	13.1	977	5	BQ882484	BQ882484	AGENCOURT
23	130	12.9	824	7	CN535522	CN535522	UI-M-HSO-
24	126	12.5	479	7	CN418585	CN418585	170004243
25	124	12.3	684	2	BE617384	BE617384	601442018
26	123	12.2	629	7	CN165820	CN165820	996655 MA
27	117	11.6	595	4	BG120637	BG120637	602346774
28	113	11.2	480	6	CB728664	CB728664	AWGNNUC:N
29	113	11.2	682	4	BI562457	BI562457	603256345
30	110	10.9	778	7	CN418584	CN418584	170004241
31	108	10.7	552	6	CB606093	CB606093	AMGNNUC:N
32	106	10.7	861	5	BU322224	BU322224	603488266
33	106	10.5	369	9	CG672107	CG672107	RRM246 Ba
34	105	10.4	740	5	BM944240	BM944240	UI-M-EHOP
35	104	10.3	521	2	BE953921	BE953921	UI-M-CC1-
36	102	10.1	742	7	CN533008	CN533008	UI-M-HOO-
37	102	10.1	854	5	BQ423127	BQ423127	AGENCOURT
38	96	9.5	501	2	AW139939	AW139939	UI-H-B11-
39	96	9.5	527	6	CB431226	CB431226	607186 MA
40	96	9.5	561	2	AW295763	AW295763	UI-H-B11-
41	96	9.5	608	2	BF394200	BF394200	UI-R-CAO-
42	91	9.0	451	6	BY599692	BY599692	BY599692
43	91	9.0	532	2	BF075268	BF075268	224165 MA
44	91	9.0	753	7	CF535194	CF535194	UI-M-GHO-
45	91	9.0	784	6	CA328666	CA328666	UI-M-FYO-

ALIGNMENTS

RESULT 1	BU508193	1167 bp	mRNA	linear	EST 12-SEP-2002
BU508193	AGENCOURT_10128476	NIH_MGC_71	Homo sapiens	cdna clone	IMAGE:6502231
LOCUS	5', mRNA sequence.				
DEFINITION	AGENCOURT_10128476	NIH_MGC_71	GI:22814426		
ACCESSION	BU508193				
VERSION	BU508193.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1167)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM14057 row: p column: 08 High quality sequence stop: 762.				
FEATURES	Location/Qualifiers				
source	1..1167				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:6502231"				

QY 223 TyrPheHisAlaGlnCysAsnPhenGlnAspGlyLeuLysAlaValGluSerLeuLys 242
 Db 314 TACTTTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAAAGCGGTGGAAGCCTCAAA 373
 QY 243 ProSerileGluThrLeuSerThrAspLeuHisThrileLysGlnAlaGlnAspGluGlu 262
 Db 374 CTTTCCATTGAACCGTCTACGGATCTTACAGGATCAACAGGCCCGCATGTAAGAA 433
 QY 263 ArgArgGlnLeuileGlnLeuArgAspPheLeuLysSerAlaLeuGlnValGluGlnLys 282
 Db 434 AGAAGGCGAGTTGATACAGCTTCGAGATATTTTGAATCCGATTCGCGATTGGAACAGAA 493
 QY 283 GluAspSerGlnileArgGlnSerThrAlaTySerLeuHisGlnProGlnGlyAsnLys 302
 Db 494 GAGGACTCCCAAAATTCGTCAGAGCACAGCTTATAGCTTATACATCAGCCTCAGGGAACAAG 553
 QY 303 GluHisGlyThrGluAArgAspGlySerLeuTyLysLysSerAspGlyLeuArgLys-Va 322
 Db 554 GAACATGGGACCGGAGCGAGCGAGCTCTACAGAAAGAGTGACGGATCCGAAN-AGT 612
 QY 322 lTTPGlnLysArgLysCysSerValLysAsnGlyPheLeuThrileSerHisGlyThrAl 342
 Db 613 GTGGCAGAAAAGGAAATGTTTCAGTAAATAATGTTTTCTGACCATATCCCATGTTACGCG 672
 QY 342 aAsnArgProAlaLysLeuAsnLeuLeuThrCysGln 355
 Db 673 TAACCGGCTCTCGAAAGCTCAACCTGCTAACCTGCCAG 712

RESULT 3
 LOCUS AL044307 556 bp mRNA linear EST 04-SEP-2003
 DEFINITION DKFZp3402129 r1 434 (synonym: htes3) Homo sapiens cDNA clone
 DKFZp3402129 5', mRNA sequence.

ACCESSION AL044307

VERSION AL044307.1 GI:5432530

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Koehrer, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

JOURNAL EST (Koehrer, et al.)

COMMENT Unpublished (1999)

Contact: MIPS

MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by BPFZ (Biomedical Research Center at the Charite,

Berlin/Germany) within the cDNA sequencing consortium of the German

Genome Project.

No sl sequence available.

This clone (DKFZp3402129) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..556

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp3402129"

/tissue_type="testis"

/dev_stages="adult"

/lab_host="DH10B"

/clone_lib="434 (synonym: htes3)"

/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

FEATURES

source

Alignment Scores:

Pred. No.: 7.25e-175

Score: 177.00

Length: 556

Matches: 177

ORGANISM

source

1..846

/organism="Homo sapiens"

High quality sequence stop: 524.

Location/Qualifiers

1..846

/organism="Homo sapiens"

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.59% Indels: 0
 DB: 1 Gaps: 0

US-09-914-042-1 (1-1006) x AL044307 (1-556)

QY 166 GluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGlyAlaGluileAlaGlu 185

Db 2 GAACAGCCCAAGCTCCATGGGATGATTCGGACTGAAATAAGCGAGCGAAATTCGCCAA 61

QY 186 GluMetGluLysGluAArgPhePheGlnLeuMetCysGluTyTrieuLeuLysVal 205

Db 62 GAGATGGAAAAGGAGAGCGCTTCTTCAGCTACAGATGTGCGAGTATCTCTGGAAGTTC 121

QY 206 AsnGluLysIleLysLysGlyValAspLeuLeuGlnAsnLeuLysLysTyrrPheHis 225

Db 122 AACGAAATCAAGATTAAGAGGAGTAGATTACTTCAGATCTGATCAATATCTTCTAT 181

QY 226 AlaGlnCysAsnPhenGlnAspGlyLeuLysAlaValGluSerLeuLysProSerile 245

Db 182 GCCCAATGCAATTTTTTTCAGATGGACTCAAGCGGTGGAAGCCTCAAACTTCCTT 241

QY 246 GluThrLeuSerThrAspLeuHisThrileLysGlnAlaGlnAspGluLysArgGln 265

Db 242 GAAACGCTGTCTACGGATCTTCACAGCATCAACAGGCCCGCATGAAAGAAAGAGGAG 301

QY 266 LeuileGlnLeuAArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSer 285

Db 302 TTGATACAGCTTCAGATATTTTGAATCCCATTCAGGTTGAACAAAGAGAGGACTCC 361

QY 286 GlnileArgGlnSerThrAlaTySerLeuHisGlnProGlnGlyAsnLysGluHisGly 305

Db 362 CAAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGAAACCAAGAACATGGG 421

QY 306 ThrGluAArgAsnGlySerLeuTyLysLysSerAspGlyIleAArgLysValTrpGlnLys 325

Db 422 ACCGAGCGGAACGCGAGCTCTACAGAAGAGTACGCGGATCCGAAAAGTGTGGCAGAA 481

QY 326 ArgLysCysSerValLysAsnGlyPheLeuThrileSerHisGlyThrAla 342

Db 482 AGGAAATGTTTCAGTTAAATAATGGTTTTCTGACCATATCCCATGTGTACCGCT 532

RESULT 4

LOCUS BUI55022 846 bp mRNA linear EST 03-SEP-2002

DEFINITION AGENCOURT_7944696 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6148570

5', mRNA sequence.

ACCESSION BUI55022

VERSION BUI55022.1 GI:22668554

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC/BCTD/DTF

CDNA Library Preparation: Life Technologies, Inc.

CNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

Clone Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/ILNL at:

http://image.llnl.gov

Plate: LLAM13480 Row: d Column: 11

High quality sequence stop: 524.

Location/Qualifiers

1..846

/organism="Homo sapiens"

FEATURES

source

Alignment Scores:

Pred. No.: 7.25e-175

Score: 177.00

Length: 556

Matches: 177

ORGANISM

source

1..846

/organism="Homo sapiens"

High quality sequence stop: 524.

Location/Qualifiers

1..846

/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6148570"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 Kb. Library constructed by Life Technologies."

ORIGIN
Alignment Scores:
Pred. No.: 4.57e-162 Length: 846
Score: 165.00 Matches: 165
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 16.40% Indels: 0
DB: 5 Gaps: 0

US-09-914-042-1 (1-1006) x BU155022 (1-846)
Qy 535 TyrAlaArgLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVal 554
Db 26 TACGACAGAGAGACGCGGATTAACCGCGGAAGCTTTCACAGTCTTTGGAGCGCGTC 85
Qy 555 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGlu 574
Db 86 AAAACGAGAGATATTTTGGATTGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAA 145
Qy 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594
Db 146 AAAATCCCACTGGCCACGACATGAGCGGATGAACGGCCCTCCACCTTGCAGTCAGA 205
Qy 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeu 614
Db 206 TCCGTGGATCGAACCCTCTCTTCACATTTAGACTTTTATTTAGTTTTCAGAACAGTGGAACTG 265
Qy 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAla 634
Db 266 GATAAACAGACAGGAAAGGACGACAGCCCTGCACTACTGCTGCTGACCGACAATGCC 325
Qy 635 GluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654
Db 326 GAGTGCCCTCAAGTGTCTCTCGGGGAAGGCTCCATCGAGATAGCAACAGAGTCAGGA 385
Qy 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuThr 674
Db 386 GAGACTCCGCTGGACATTTCCCAAGCGCTCAAGCACGAGCACTGTGAGGAGCTGCTGACC 445
Qy 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeu 694
Db 446 CAAGCCTTATCTGGAAGATTTAATTTCTCACGTTTTCACGTTTGAATGAATGCGACTACTC 505
Qy 695 HisGluAspLeuAsp 699
Db 506 CACGAAGACCTGGAT 520

RESULT 5
BU703510
LOCUS
DEFINITION UI-M-F00-bzo-m-08-0-UI.r1 NIH_BMAP_F00 Mus musculus CDNA clone
IMAGE:6405295 5', mRNA sequence.
ACCESSION BU703510.1 GI:23629426
VERSION BU703510.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.
Location/Qualifiers
FEATURES
Source
1..689
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6405295"
/tissue_type="whole brain"
/lab_host="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_F00"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGAGCC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Alignment Scores:
Pred. No.: 6.83e-157 Length: 689
Score: 160.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 15.90% Indels: 2
DB: 5 Gaps: 0

US-09-914-042-1 (1-1006) x BU703510 (1-689)
Qy 312 LeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331
Db 99 CTCTACAAGAAGACGATGGATCCGGAAGTGTGGCAGAGAGAGAAAGTGTTCGGTTAA 158
Qy 332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351
Db 159 AATGGCTTCTTCACCATCTCCAGGCACCTGCCAACCGGCCCTGCCAAGCTCAACCTG 218
Qy 352 LeuThrCysGlnValLysThrAsnProGluLysLysCysPheAspLeuIleSerHis 371
Db 219 CTAACTGCCAGGTGAAGACCAATCTCGAGAGAAGTGTTCGACCTCATATCAT 278
Qy 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTyrMetSerVal 391
Db 279 GACAGGACGTACCACTTCCAAAGACGACCAAGAAATGTCAGATATGATGTCGTGA 338
Qy 392 LeuGlnAsnSerLysGluAlaLeuAsnAlaPheLysGlyAspAspAsnThrGly 411
Db 339 CTGCAGAACACCAAGGAAGAGCTCTGAACACGCCCTTTAAGGGTGTATGACACACTGGA 398
Qy 412 GluAsnAsnIleValGlnGluLeuThrLysGluIleSerGluValGlnArgMetThr 431

```

```

Db      399 GAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGGAGTGCAGAGGATGACG 458
Qy      432 GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeu 451
Db      459 GGCATACAGCTGTGCTGCGACTGTGGGCGACCCAGATCCGAGCTGCTCTTACCAACCTG 518
Qy      452 GlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTrpSer 471
Db      519 GGCATCCCTGACTTGCATCGAGTCTCTGGGATCCACCGGAGCTGGGGGTTTCATTACTCC 578
Qy      472 Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAs 491
Db      579 AG-GATGAGTGCCTGACATTAGATGATGGGAACGCTGAGCTTCCTGCTGCCAGAA 637
Qy      491 nileGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuPro 506
Db      638 TATCGGAATGCAGGCTTTAATGAAATTATGGAGTGTTCCTACCG 683

RESULT 6
CN538273      731 bp      mRNA      linear      EST 29-APR-2004
LOCUS      UI-M-H50-cqr-k-12-0-UI.r1 NIH_BMAP_H50 Mus musculus cDNA clone
DEFINITION      IMAGE:30675611 5', mRNA sequence.
ACCESSION      CN538273
VERSION      CN538273.1 GI:46866429
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
FEATURES      Location/Qualifiers
source      1..731
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="IMAGE:30675611"
            /tissue_type="Upper Head"
            /dev_stages="embryo 9.5 - 10.5 dpc"
            /lab_host="DH10B (T1 phage resistant)"
            /clone_lib="NIH_BMAP_H50"
            /note="Organ: Upper Head; Vector: pYX-Asc; Site 1: Ecor I;
            Site 2: Not I; The library was constructed according
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with Ecor I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is CCAATGAAT. This library was created for the University
            Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH)".
ORIGIN
Alignment Scores:

```

```

Pred. No.:      7,23e-157      Length:      731
Score:      160.00      Matches:      192
Percent Similarity:      98.97%      Conservative:      0
Best Local Similarity:      98.97%      Mismatches:      1
Query Match:      15.90%      Indels:      2
DB:      7      Gaps:      0

US-09-914-042-1 (1-1006) x CN538273 (1-731)

Qy      312 LeuTrpLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331
Db      152 CTCTACAGAGAGCGATGGATCCGGAAGTGTGGCAGAGAGAGAAAGTGTTCGTTAAA 211
Qy      332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeu 351
Db      212 AATGGCTTCCTCACCATCTCCACGGCACTGCCAACCGCCACCTGCCAAGCTCAACCTG 271
Qy      352 LeuThrCysGlnValLysThrAsnProGluLysLysCysPheAspLeuIleSerHis 371
Db      272 CTAACTCTGCCAGGTGAAGACCAATCTCTGAGAGAGAAAGTGTTCGACCTCATATCACAT 331
Qy      372 AspArgThrTrpHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerVal 391
Db      332 GACAGGACGTACCACTTCCAAAGCGGAGACGACGACGAAATGTTCAGATATGATGTCTGTA 391
Qy      392 LeuGlnAsnSerLysGluAlaLeuAsnAlaPheLysGlyAspAsnThrGly 411
Db      392 CTGCAGAACCAAGGAAGAGCTCTGAACAACGCCTTTAAGGGTGTATGACCAACACTGGA 451
Qy      412 GluAsnAsnIleValGlnGluLeuThrLysGluIleSerGluValGlnArgMetThr 431
Db      452 GAAATAACATCGTCCAGAGCTGACCAAGAGAGATCATCTCGGAGGTGCAGAGGATGACG 511
Qy      432 GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeu 451
Db      512 GGCATGACGTGTCTGCGACTGTGGGCGACACAGATCCGAGCTGTCTTACCAACCTG 571
Qy      452 GlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTrpSer 471
Db      572 GGCATCTCTGACTTGCATCGAGTGTCTCGGATTACCGGGAGCTGGGGGTTTCATTACTCC 631
Qy      472 Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs 491
Db      632 AG-GATGAGTCCCTGACATTAGATGTATTGGGAACGCTCGAGCTTCTGCTTGCACAGAA 690
Qy      491 nileGlyAsnAlaGlyPheAsnGluIleMetGluCysCys 504
Db      691 TATCGGAATGCAGGCTTTAATGAAATTATGGAGTGTTCG 730

RESULT 7
CN460643      773 bp      mRNA      linear      EST 21-APR-2004
LOCUS      UI-M-HB0-coy-a-05-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
DEFINITION      IMAGE:30652708 5', mRNA sequence.
ACCESSION      CN460643
VERSION      CN460643.1 GI:46466369
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html

```

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1. 773
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30652708"
/issue_type="whole eye"
/lab_host="embryo 12.5.13.5,14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HB0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTATTGAAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
Pred. No.: 7,63e-157 Length: 773
Score: 160.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 15.90% Indels: 2
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CM460643 (1-773)

Qy 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerVallys 331
Db 53 CTCTACAGAGAGCGGATGGGATCGGAAGTGTGGCAGAGAGAGAGTTCGGTTAA 112
Qy 332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351
Db 113 AATGGCTTCTCACCATCTCCACGGCACTGCCAACCGGCCACCTGCCAAGCTCAACCTG 172
Qy 352 LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371
Db 173 CTACCTGCCAGGTGAGACCACTCTGAGGAGAGAGAGTGTTCGACCTCATATCAT 232
Qy 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerVal 391
Db 233 GACAGGAGTACCCTTCCAGCGGAGAGCAACGAAGATGTCAGATGATGATGTCGTGA 292
Qy 392 LeuGlnAsnSerLysGluLysAlaLeuAsnAlaPheLysGlyAspAsnThrGly 411
Db 293 CTGCAGAACAGCAAGAGAGAGCTCTGAACACGCTTTAAAGGGTGATGACACACTGGA 352
Qy 412 GluAsnAsnIleValGlnLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
Db 353 GAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGAGGTGCAGAGGATGACG 412
Qy 432 GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeu 451
Db 413 GGCAATGACGTGTCTGCGACTGTGGGCGCACCAGATCCGAGTGGCTCTCTACCAACCTG 472
Qy 452 GlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSer 471
Db 473 GGCATCCTGACTGTGATGCTCTGGGATCCACCGGGAGCTGGGGGTTCATTACTCC 532
Qy 472 Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs 491

Db 533 AG-CATCAGTCCCTGACATTAGATGATTATGGGACGCTCTGAGCTTCTGTTGCCAGAA 591
Qy 491 nileGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuPro 506
Db 592 TATCGGAATGCAGGCTTTAATGAATATATGAGTGTTCCTACCG 637

RESULT 8

CO429713
LOCUS
DEFINITION
UI-M-HX0-crw-p-02-0-UI.r1 NIH_BMAP_HX0 Mus musculus cDNA clone
IMAGE:30686473 5', mRNA sequence.
CO429713
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
REFERENCE
1 (bases 1 to 781)
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaba-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1. 781
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30686473"
/issue_type="whole eye"
/tissue_type="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HX0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATAATACG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
Pred. No.: 7,71e-157 Length: 781
Score: 160.00 Matches: 194
Percent Similarity: 98.98% Conservative: 0
Best Local Similarity: 98.98% Mismatches: 1
Query Match: 15.90% Indels: 2
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CO429713 (1-781)

Qy 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerVallys 331

```

Db      83  CTCTCAAGAGAGGCGATCGGATCCGAAAGTGTGCAGAGAGGAAGTGTCCGTTAAA 142
Qy      332  AaNGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351
Db      143  AATGCTTCTCCATCCATCTCCACGCGACTGCGCAACCGGCCACCTGCCAAGCTCAACCTG 202
Qy      352  LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371
Db      203  CTAACCTGCCAGGTGAAGACCAATCTCTGAGGAGAGAAAGTGTTCGACCTCATATCACAT 262
Qy      372  AspArgThrTyHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerVal 391
Db      263  GACAGAGTACCATCTCAAGCGGAGACGAAAGAAATGTCAGATGATGATGTCTGTA 322
Qy      392  LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGly 411
Db      323  CTGCAGAACAGCAAGGAAGAGCTCTGAACACGCGCTTTAAGGTGTGATGACACACTGGA 382
Qy      412  GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
Db      383  GAAATAAATCATCGTCAAGAGCTGACCAAGGAGATCATCTCGAGGTGCAGAGGATGACG 442
Qy      432  GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeu 451
Db      443  GCAATAGCTGTGTCTGCGACTGTGGGCGACGATCCGACGTGCTCTTACCAACCTG 502
Qy      452  GlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTySer 471
Db      503  GGCATCTGCTGCTGATCGAGTCTCTGGATCCACCGGAGCTGGGGTTCATTACTCC 562
Qy      472  Pro-MetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs 491
Db      563  AG-GATGCAGTCCCTCGACATTAGATGTATTGGAAAGCTCTGAGTCTCTGCTTGTCCAGAA 621
Qy      491  nileGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuPro 506
Db      622  TATCGGGAATGCGAGCTTAAATGAATAATTATGGAGTGTTCGCTTACCG 667

```

```

RESULT 9
BP280849 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
LOCUS
DEFINITION
BP280849 KGR09051, mRNA sequence.
ACCESSION
BP280849
VERSION
BP280849.1 GI:52194581
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Suzuki, Y., Yamaehita, R., Shiota, M., Sakakibara, Y., Chiba, J.,
Mizushima-Sugano, J., Nakai, K. and Sugano, S.
Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
Genome Res. 14 (9), 1711-1718 (2004)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp.
Location/Qualifiers
1. .582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KGR09051"
/cell_line="KG-1-C"
/clone_lib="Sugano cDNA library, KG-1-C"
/note="Glioma"

```

FEATURES

```

source
1. .582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KGR09051"
/cell_line="KG-1-C"
/clone_lib="Sugano cDNA library, KG-1-C"
/note="Glioma"

```

ORIGIN

```

source
1. .485
/organism="Homo sapiens"
/mol_type="mRNA"

```

```

Alignment Scores: .
Pred. No.:      8.32e-154      Length:      582
Score:          157.00      Matches:      157
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatch: 0
Query Match:      15.61%      Indels:      0
Db:              5      Gaps:      0

```

US-09-914-042-1 (1-1006) x BP280849 (1-582)

```

Qy      43  AlalaLeuAspValAspArgMetValLeuTyLysMetLysLysSerValLysAlaIleAan 62
Db      112  GCTTTGGAGCTGGACCGGATGTTCTTTACAAATGAAGAAATCCGTGAAGCAATCAAC 171
Qy      63  SerSerGlyLeuAlaHisValGluAsnGluGlnTyThrGlnAlaLeuGluLysPhe 82
Db      172  AGCTCTGGCTGGCTCACGTGAAATGAAGAGCAGTACACCCAGGCTCTGGAGAAGTTT 231
Qy      83  GlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLysPheSer 102
Db      232  GCGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTGGCTTCTCGAAGTTCTCA 291
Qy      103  ValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsnIle 122
Db      292  GTGTTTCAAAAGGAGTTGACAGCACTTTTCAAAACCTGATTTCAGATATGAACACATA 351
Qy      123  IleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGlyAspLeu 142
Db      352  ATCTCTTCCCTTTGGACAGTTTCTGAAGGGGACCTGAAAGGAGTGAAGGGGATCTG 411
Qy      143  LysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIleGluLys 162
Db      412  AAAAAGCCTTTTGTATAAGCTTGGAAAGACTATGAAACAAATAAACAAGATAGAAAG 471
Qy      163  GluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGlu 182
Db      472  GAGAAAAAGGAACACGCCAAGCTTCATGGGATGATTTCGACTGAAATGAAGCGGCGAA 531
Qy      183  IleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCys 199
Db      532  ATTCCGAGAGATGGAAAGAGAGGCGCTTCTTCAGCTACAGATGTGC 582

```

```

RESULT 10
BM839307
LOCUS
DEFINITION
K-EST0116049 S9SNU601 Homo sapiens cDNA clone S9SNU601-68-A09 5',
BM839307 485 bp mRNA linear EST 06-MAR-2002
mRNA sequence.
ACCESSION
BM839307
VERSION
BM839307.1 GI:19195716
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 485)
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 68 row: A column: 09
High quality sequence stop: 485.
Location/Qualifiers
1. .485
/organism="Homo sapiens"
/mol_type="mRNA"

```

```

ACCESSION
BM839307
VERSION
BM839307.1
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

```

```

REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
COMMENT

```

```

FEATURES
source
1. .485
/organism="Homo sapiens"
/mol_type="mRNA"

```



```

Db      183  CAGCGGAAGACCAACAAGATGTCAGATATGATGATGCTGTACTCCAGAACAGCAAGGAA 242
Qy      398  GluAlaLeuAsnAlaPheLysGlyAspAsnThrGlyGluAsnAlaValGln 417
Db      243  GAAGCTCTGAACAACAGCGCTTTAAAGGGTGATGACACACATGGAGAAAATAACATCGTCCAA 302
Qy      418  GluLeuThrLysGluLeuIleSerGluValGlnArgMetThrGlyAsnAspValCys 437
Db      303  GAGCTGACCAAGGAGCATCATCTCGGAGGTGACAGGATGACGGGCAATGACGTGTGCTGC 362
Qy      438  AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyLeuThrCysIle 457
Db      363  GACTGTGGGGACCAAGATCCGACGTGGCTCTCTACCAACCTGGGCATCTGACTTGGCATC 422
Qy      458  GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuTh 477
Db      423  GAGTCTCTGGGATCCACGGGAGCTGGGGGTTCATTACTCCAG-GATGCAGTCCCTGAC 481
Qy      477  rleuaspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPh 497
Db      482  ATTAGATGATTGGGAAGCTCTGAGCTTCTGCTTCCCAAGATATCGGGAATGCAAGGCTT 541
Qy      497  eAsnGluIleMetGluCysCysLeuPro 506
Db      542  TAATGAAATTATGGAGTGTTCCTACCG 569

RESULT 12
LOCUS   BQ439046 884 bp mRNA linear EST 24-MAY-2002
DEFINITION AGENCOURT_7761529 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020058
5', mRNA sequence.
ACCESSION BQ439046
VERSION   BQ439046.1 GI:21178122
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgi.nci.nih.gov/
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13223 row: i column: 19
High quality sequence stop: 733.
Location/Qualifiers
1. .884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6020058"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

FEATURES
source
1. .884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6020058"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

ORIGIN
Alignment Scores:
Pred. No.: 1.78e-150 Length: 884
Score: 154.00 Matches: 242
Percent Similarity: 99.18% Conservative: 0
Best Local Similarity: 99.18% Mismatches: 1

```

```

Query Match: 15.31% Indels: 2
DB:          5          Gaps: 0
US-09-914-042-1 (1-1006) x BQ439046 (1-884)
Qy      318  GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db      13  GGGATCCGAAGAAGTGTGGCAGAAAAGGAATGTTTCAGTTAAAAATGGTTTCTGACCATATA 72
Qy      338  SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db      73  TCCCATGTGTACCGCTAACCGGCTCTCTGCAAGGCTCAACCTGCTAACCTGCCAGGTGAAG 132
Qy      358  ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
Db      133  ACCAACCTGAGGAGAGAGAGTGTGACCTTATTTTACATGACAGAACTTACCACCTTT 192
Qy      378  GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGluAsnSerLysGlu 397
Db      193  CAAGCTGAAGATGAACAGGAATGTCAAATATGGATGCTGTCTGCTGCAAAATAGCAAGAA 252
Qy      398  GluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAlaValGln 417
Db      253  GAAGCTTTAAACAATGCATTTAAGGGGATGACAACTACTGGAGAAAATAACATCGTCCAA 312
Qy      418  GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
Db      313  GAACCTGACAAAGGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGCTGCTGT 372
Qy      438  AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457
Db      373  GACTGTGGGGCCGACAGATCCTTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATC 432
Qy      458  GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuTh 477
Db      433  GAGTGTTCGGGAATCCACCGAGAGCTGGGGGTTCATTATTTCAG-GATGCAGTCCCTGAC 491
Qy      477  rLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPh 497
Db      492  CTTAGATGATCTGGGACATCTGAGCTGCTGCTCGCCAAGAATATTGGGAATGCAAGCTT 551
Qy      497  eAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySe 517
Db      552  TAATGAGATCATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCAACCCAGGCGAG 611
Qy      517  rAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTgTyrAlaAr 537
Db      612  CGCATGAATGCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAG 671
Qy      537  GlyLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuGlyGluAlaValLysThrAr 557
Db      672  GAAGAAGCACCGGATTAACGCGCGGAGCTTACAGTCTTTTGGGAGGCGGTCAAAACCGAG 731
Qy      557  gAspIlePhe 560
Db      732  AGATATTTTT 741

RESULT 13
LOCUS   CD358314 948 bp mRNA linear EST 29-MAY-2003
DEFINITION AGENCOURT_14275993 NIH_MGC_180 Homo sapiens cDNA clone
IMAGE:30388565 5', mRNA sequence.
ACCESSION CD358314
VERSION   CD358314.1 GI:31129749
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgi.nci.nih.gov/
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```

COMMENT

Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: NDAM460 row: k column: 06
High quality sequence stop: 543.
Location/Qualifiers
1. .948
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30388565"
/lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
/note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dr primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

FEATURES

source

ORIGIN

Alignment Scores:
Pred. No.: 2,738-147 Length: 948
Score: 151.00 Matches: 179
Percent Similarity: 98.90% Conservativeness: 0
Best Local Similarity: 98.90% Mismatches: 1
Query Match: 15.01% Indels: 2
DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x CD358314 (1-948)

QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
DB 3 AAAGTGTGCAGAAAAGGAATGTTTCAGTTAAATAGTTTCTGCACCATATCCCATGTT 62
QY 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
DB 63 ACCGCTAACCGGCCCTCTGCAAAAGTCAACCTGCTAACCTGCAGGTGAAGACCAACCTT 122
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
DB 123 GAGGAGAGAAGTGTGTTGACCTTATTTCCATGATGACAGACTTACCATCTTCAAGCTGAA 182
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
DB 183 GATGAACAGGAATGTCAATATAGGATGCTGTGCTGCAAAATAGCAAGAAAGAGCTTTA 242
QY 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
DB 243 AACAAATGCAATTTAAAGGGGATGACAATACTGGAGAAAATAAATCATGCTCCAAAGAACTGACA 302
QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
DB 303 AAGGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGTCTGTGCTGACGTGGG 362
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
DB 363 GCGCAGATCCTACATGGCTTCCACCAACCTGGGCATCCTGACCTGCATCAGGTGTTCC 422
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspVa 480
DB 423 GGAAATCCACCGAGAGCTGGGGGTTTCAATTATTCAG-GATGCAGTCCCTGACCTTAGATGT 481
QY 480 lleuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500

Db 482 ACTGGGAACATCTGAGCTGCTGCTCGCAAGAAATATGGGAATGAGCTTTAATGAGAT 541
QY 500 e 500
DB 542 C 542
RESULT 14
CF745192
LOCUS
DEFINITION
CF745192
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CF745192 714 bp mRNA linear EST 10-OCT-2003
UI-M-GVO-clv-g-03-0-UI.r1 NIH_BMAP_GVO Mus musculus cDNA clone
IMAGE:30620594 5', mRNA sequence.
CF745192
CF745192.1 GI:37641532
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
seq primer: pYX-5.
Location/Qualifiers
1. 714
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30620594"
/tissue_type="whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GVO"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated with
EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAACTGAAT. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 2,986-144 Length: 714
Score: 148.00 Matches: 148
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.71% Indels: 0
DB: 7 Gaps: 0
US-09-914-042-1 (1-1006) x CF745192 (1-714)

QY 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331

```

Db      90  CTCTACAGAGAGGAGGATCGGATCCGAAAGTGTGCGAGAGAGAGAGTTCCTTTAAA 149
Qy      332  AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeu 351
Db      150  AATGCTTCTTCAACCATCTCCACAGGCACTGCCAACCGGCCACCTGCCCAAGCTCAACCTG 209
Qy      352  LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371
Db      210  CTAACCTCCAGGTGAAGACCAATCTCGAGGAGAGAAAGTGTTCGACCTCATATCACAT 269
Qy      372  AspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTyrMetSerVal 391
Db      270  GACAGAGCTACCATCTCCAGCGGAGACCAAGAAATGTCAGATGATGATGCTGTA 329
Qy      392  LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGly 411
Db      330  CTGCAGAACACAGCAAGAAAGACTCTGAACACGCTTTAAGGTGATGACACACTGA 389
Qy      412  GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
Db      390  GAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGAGGTGCAGAGTACG 449
Qy      432  GlyAsnAspValCysCysAspCysGlyAlaProAspProThrTyrLeuSerThrAsnLeu 451
Db      450  GGCAATGACGTGTGCTGCGACTGTGGGCGCACAGATCCGAGCTGCTCTACCAACCTG 509
Qy      452  GlyIleLeuThrCysIleGluCys 459
Db      510  GGCATCTGACTTGCATCGATGTC 533

```

```

RESULT 15
LOCUS   CN535077
DEFINITION UI-M-HS0-cqi-a-11-0-UI-r1 NIH_BMAP_H50 Mus musculus cDNA clone
IMAGE:30674218 5', mRNA sequence.
ACCESSION CN535077
VERSION   1
KEYWORDS  Mus musculus (house mouse)
SOURCE    Mus musculus
ORGANISM  Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 742)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefi.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5
Location/Qualifiers
1..742
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30674218"
/tissue_type="Upper Head"
/dev_stage="embryo 9.5 - 10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_H50"
/note="Organ: Upper Head; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose

```

FEATURES

```

source
1..742
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30674218"
/tissue_type="Upper Head"
/dev_stage="embryo 9.5 - 10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_H50"
/note="Organ: Upper Head; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose

```

gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

```

Alignment Scores:
Pred. No.: 3,09e-144 Length: 742
Score: 148.00 Matches: 182
Percent Similarity: 98.91% Conservative: 0
Best Local Similarity: 98.91% Mismatches: 1
Query Match: 14.71% Indels: 2
DB: 7 Gaps: 0

```

US-09-914-042-1 (1-1006) x CN535077 (1-742)

```

Qy      324  GlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsn 343
Db      69  CAGAAGAGAGAGTGTTCGTTAAATAATGGCTTCTCCACATCTCCACGGCACTGCCAAC 128
Qy      344  ArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnProGluGluLys 363
Db      129  CGGCCACCTGCCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAATCTTGAGGAGAAG 188
Qy      364  LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383
Db      189  AAGTGTTCGACCTCATATCATCATGACAGGACGTACCACTTCCAGCGGAGAGACGACAA 248
Qy      384  GluCysGlnIleTyrMetSerValLysGlnAsnSerLysGluGluAlaLeuAsnAla 403
Db      249  GAATGTCAGATATGATGTCTGTACTGCAAGAACAGCAAGGAGAGAGCTCTGAAACACGCC 308
Qy      404  PheLysGlyAspAspAsnThrGlyLysAsnIleValGlnGluLeuThrLysGluIle 423
Db      309  TTTAAGGTGATGACACACACTGGAGAAATAAATCAATCGTCCAAGAGCTGACCAAGGAGATC 368
Qy      424  IleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAsp 443
Db      369  ATCTCGAGGTGTCAGAGGATGACGGCAATGACGTGTCTGCGACTGTGGGCGCACCAT 428
Qy      444  ProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHis 463
Db      429  CCGACGTGGCTCTCTTACCACCTGGGCATCTGACCTTGCATCGAGTGTCTTGGGATCCAC 488
Qy      464  ArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspValLeuGlyTh 483
Db      489  CGGAGCTGGGGGTTCATTACTCCAG-GATCGAGTCTCCATAGATAGATGATTGGGAAC 547
Qy      483  rSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCys 503
Db      548  GTCTGAGCTTCTGCTTGCAGAAATATCGGAATATCGGCAATCGAGCTTTAATGAATATGAGGTG 607
Qy      503  sCysLeuPro 506
Db      608  TTGCTTACCG 617

```

Search completed: August 5, 2005, 00:18:10
Job time : 7514 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 22:16:06 ; Search time 1433 Seconds
(without alignments)
4550.746 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHEDYKA.....DGDGPKRGKAPVPSFVHFIAD 1006

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7297361 seqs, 3241162794 residues

Word size: 1

Total number of hits satisfying chosen parameters: 14583226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+p2n.model -DB=FASTA
-Q=/cgn2_1/USPTO_spool_h/US09914042/runat_04082005_090749_2764/app_query.fasta_1.1159
-DB=FASTA -Q=/cgn2_1/USPTO_spool_h/US09914042/runat_04082005_090749_2764/app_query.fasta_1.1159
-LOOPEXT=0 -LOOPEXT=0 -LOOPEXT=0 -LOOPEXT=0 -LOOPEXT=0 -LOOPEXT=0 -LOOPEXT=0 -LOOPEXT=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09914042 @CCN 1 1 658 @runat 04082005_090749_2764
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	717	71.3	5711	21	US-10-956-1157-1144
2	110	10.9	3123	17	US-10-104-047-314
3	110	10.9	3436	11	US-09-764-875-37
4	79	7.9	3413	11	US-09-764-875-358
5	78	7.8	442	10	US-09-318-995-27010
6	48	4.8	563	9	US-09-815-343-1154
7	48	4.8	563	18	US-10-097-105-1154
8	38	3.8	958	17	US-10-027-632-34151
9	38	3.8	958	17	US-10-027-632-34151
10	30	3.0	409	18	US-10-276-774-404
11	30	3.0	3974	18	US-10-210-281-73
12	22	2.2	5582	17	US-10-062-674-1606
13	19	1.9	60	10	US-09-908-975-22014
14	17	1.7	326	9	US-09-815-343-1151
15	17	1.7	326	18	US-10-097-105-1151
16	16	1.6	469	18	US-10-276-774-288
17	16	1.6	1523	21	US-10-764-420-1028
18	14	1.4	1038	18	US-10-276-774-639
19	13	1.3	167	9	US-09-864-761-30407
20	13	1.3	323	9	US-09-728-446-1051
21	13	1.3	446	9	US-09-864-761-13841
22	12	1.2	296	20	US-10-333-872A-161
23	12	1.2	399	17	US-10-442-535A-21815
24	12	1.2	399	18	US-10-085-783A-21815
25	12	1.2	4211	17	US-10-172-118-1101
26	12	1.2	4211	18	US-10-342-887-1101
27	12	1.2	28055	17	US-10-085-117-94
28	11	1.1	451	17	US-10-062-674-482
29	10	1.0	417	9	US-09-912-020-140
30	10	1.0	417	20	US-10-771-241-140
31	10	1.0	697	9	US-09-912-020-446
32	10	1.0	697	20	US-10-771-241-446
33	10	1.0	824	13	US-10-027-632-157073
34	10	1.0	824	17	US-10-027-632-157073
35	10	1.0	2339	13	US-10-027-632-110268
36	10	1.0	2339	13	US-10-027-632-110268
37	10	1.0	2339	13	US-10-027-632-110268
38	10	1.0	2339	17	US-10-027-632-110268
39	10	1.0	2339	17	US-10-027-632-110268
40	10	1.0	2339	17	US-10-027-632-110268
41	10	1.0	3812	17	US-10-104-047-981
42	10	1.0	4050	21	US-10-490-605-1
43	9	0.9	201	19	US-10-741-601-21882
44	9	0.9	307	21	US-10-487-561-51
45	9	0.9	317	19	US-10-437-963-21619

ALIGNMENTS

RESULT 1
US-10-956-157-1144
; Sequence 1144, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956.157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1144
; LENGTH: 5711

; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-956-157-1144			
Alignment Scores:			
Pred. No.:	0	Length:	5711
Score:	717.00	Matches:	1003
Percent Similarity:	99.41%	Conservative:	0
Best Local Similarity:	99.41%	Mismatches:	3
Query Match:	71.27%	Indels:	6
DB:	21	Gaps:	0
US-09-914-042-1 (1-1006) x US-10-956-157-1144 (1-5711)			
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla	20
Db	341	ATGCGGACACAGATCTCGGTGCGAATTCGTGGCCGAGACCCATGAGGACTACAGGCG	400
Qy	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaIle	40
Db	401	CCCACGGCTCCAGCTTACACCCGCGCAGTGCCGGAACACTGTGGCGGCCATC	460
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla	60
Db	461	GAGGAGGCTTTGGACGTGGACCGGATGGTTCTTTACAAAATGAAGAAATCCGTGAAGCA	520
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu	80
Db	521	ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGAGCAGTACACCCAGGCTCTGGAG	580
Qy	81	LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys	100
Db	581	AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTGGCTTCTCGAAG	640
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	641	TTCTCAGTGTTTACAAAGGAGTTGACAGCACTTTTCAAAAACCTGATTTCAGAAATATGAAC	700
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	701	AACATAATCTCTTCCCTTTGGACAGTTTGTCTGAAGGGGACCTGAAAGGAGTGAAGGG	760
Qy	141	AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle	160
Db	761	GATCTGAAAAGCCCTTTTGATAAAGCTTGAAGGACTATGAAACAAAATAACCAAGATA	820
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
Db	821	GAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATAAGCGGA	880
Qy	181	AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu	200
Db	881	CGCGAAATTCGCGAAGAGATGGAANAAGAGAGGCGCTTCTTCCAGCTACAGATGTGCGAG	940
Qy	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	941	TATCTGTGAGGTCACAGAAATCAAGATTAAAGAGGAGTAGATTACTTCAGAAATCTG	1000
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer	240
Db	1001	ATCAAAATCTTTCATGCCCAATGCAATTTTTTTTTCAGGATGGACTCAAAAGCGGTGAAAGC	1060
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1061	CTCAAACTCTCATTTGAAACGCTGTCTACGGATCTTTCACAGATCAAAACAGGCCCCAGGAT	1120
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	1121	GAAGAAAGAGGCAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCGAGGTTGAA	1180
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1181	CAGAAAGAGGACTCCCAAAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGA	1240

Qy	301	AsnLysGluHisGlyThrGluAtqAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1241	AACAAGGAACATGGGACCGAGCGGACGCGCTCTACAGAAGAGTGCAGGATCCGA	1300
Qy	321	LysValTrpGlnLysArgLysCysSerValLysShenGlyPheLeuThrIleSerHisGly	340
Db	1301	AAAGTGTGGCAGAAAAGGAAATGTTCAAGTTAAAAATGGTTTCTGACCATATCCATGGT	1360
Qy	341	ThrAlaSerArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1361	ACCGCTAACCGGCTCTCTGCAAAAGCTCAACGTGAACCTGCCAGGTGAAGACCAACCT	1420
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1421	GAGGAGAAGAAGTCTTTTGACCTTATTTTACATGACAGAACTTACCACTTTCAAGCTGAA	1480
Qy	381	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400
Db	1481	GATGAACAGGAATGTCAAAATATGATGTCTGTGCTGCAAAATAGCAAAAGAAAGCTTTA	1540
Qy	401	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1541	AACAATGCATTTTAAGGGGATGACAAATCTGGAGAAAATAACATCGTCCAGAACTGACA	1600
Qy	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1601	AAGGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGTCTGCTGTGACTGTGGG	1660
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1661	GCGCCAGATCTACATGGCTTTTCCACCAACCTGGGCATCTCGACCTCGAGTGTTC	1720
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspVa	480
Db	1721	GGAAATCCACCAGAGAGCTGGGGGTTTCATTTATTCACAG-GATGCAGTCCCTGACCTTAGATGT	1779
Qy	480	LLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1780	ACTGGGAACATCTGAGCTGTCTGCCAAGAAATATTGGGNAATCGAGGCTTTAATGAGAT	1839
Qy	500	eMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAs	520
Db	1840	CATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCAAACCCAGCAGCAGCATGAA	1899
Qy	520	nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHi	540
Db	1900	TGCAAGAAAGGACTACATCACAGCCAAATACATCGAGAGGAGATACGCAAGAAAGAGCA	1959
Qy	540	sAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh	560
Db	1960	CGCGGATAACCGCGGAGCTTTCAGCTCTTTGCGAGGCGCTCAAAAACGAGAGATATTTT	2019
Qy	560	eGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAs	580
Db	2020	TGGATTGCTCCCAAGCTTATGCTGATGTGTGATCTTTACGGAAAAAATCCCACTGGCCAA	2079
Qy	580	nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe	600
Db	2080	CGGACATGAGCGGATGAAACGGCCCTCCACTTTCAGTTCAGATCCGCTGGATCGAAACCTC	2139
Qy	600	rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy	620
Db	2140	TCTTCACATTTGATGATCTTTTAGTTTCAAGACAGTGGGACCTGGATAAACAGACAGCGAA	2199
Qy	620	sGlySerThrAlaLeuHisTyrCysLysLeuThrAspAsnAlaGluCysLeuLysLeuLe	640
Db	2200	AGGCAGCACAGCCCTCGCACTACTGCTGCTGACCGCAATGCGGAGTGCCTCAAGTTGCT	2259
Qy	640	uLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle	660
Db	2260	CCTCGGGGGAGGCTCCATTCGAGATAGCAAAACGAGTCAGGAGAGACTCCCTGGACAT	2319

```
Qy 660 eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr 580
Db 2320 TGCCAAAGCGCTCAAGACGAGAGCTGTGAGAGAGCTGTGACCCCAAGCCCTTATCTGGAG 2379
Qy 680 gPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspG1 700
Db 2380 ATTAAATCTCACGTTACGTTGATATGATGATGGGACTATCTCCAGGAAGACCTTGGATGA 2439
Qy 700 uSerAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArgP 720
Db 2440 AAGTATGATCACGATGATGATGAGAAATTCAGCCAGTCC-CAACCGCGGGAAGACCGGC 2498
Qy 720 rolleSerPheTyrClnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerleuAlaA 740
Db 2499 CCATCAGCTTTACAGCTGGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCA 2558
Qy 740 rGAspAlaAAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerileLeuGlnA 760
Db 2559 GAGATGCTGCAACCTTCCAGAGGAAGACGAGGGCTTTTCATGCCAGCATCTTGCAGA 2618
Qy 760 snGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProS 780
Db 2619 ATGAGACTTACGAGCCCTCTGAGTGGCAGCCACCTCCGCCAGCCTGCAGCCGCCA 2678
Qy 780 erThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerS 800
Db 2679 GCACCCAGCGCCCCCGCTTCCCTCACGGAATTTGGCAAAAGTTTCAGACAGCCTCCT 2738
Qy 800 erAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS 820
Db 2739 CTGCTTAACACCTGTGGAAGACAACCTCTGAAGTGTGGAGCGGTGGAAGCCGGCAGCGAT 2798
Qy 820 erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA 840
Db 2799 CTTCGTGATGATCCGACCTGCTCCATCCACGCTGCCCTCTTCGCGTGACATCTACCA 2858
Qy 840 snProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaL 860
Db 2859 ATCCCTTGACCCCAAGCGCGCCGCCACCGTTGCCAAGACGCCACCGTAATGGAAGCCT 2918
Qy 860 euSerGlnProSerLysProAlaProProGlyLysLeuSerGlnLysArgProProLeuP 880
Db 2919 TGAGCCAGCCGAGCAAGCTCCCGCGCTGGGATCTTCAGATCAGGCCCCCACCCTCTGC 2978
Qy 880 roProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer 899
Db 2979 CCCCACAGCCGCCACCGCCCTCCCGCAGAGAAGCCTGCCCGGGGCT-GACAAAGTCC 3037
Qy 900 ThrProLeuThrAsnLysGlyGlnProProArgGlyProValAspLeuSerAlaThrGluAla 919
Db 3038 ACCCAGCTGACCAACAAAGGCCAACCGAGAGGAGCTGTGGATCTCTCTGCAACGGAAGCT 3097
Qy 920 LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer 939
Db 3098 CTGGGCTCTGTCTCAATGCTATGGTCTCCAGCGCCCTGCAACCCATGCTCTAGGAAGTCG 3157
Qy 940 GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsn 959
Db 3158 CAGGCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATAACTGTGTGGCTGACAAAC 3217
Qy 960 ProAspGluLeuThrPheSerGluGlyAspVallelleValAspGlyGluGluAspGln 979
Db 3218 CCCGATGAGCTCACCTTCTCCAGGGGATGTGATCATCTGTCGAGCGGGGAGGAGCACAG 3277
Qy 980 GluTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer 999
Db 3278 GAGTGGTGGATTGGCCACATTGATGGAGATCTGTGTCGAAAGGCGCATTCCTCCGCTGTCA 3337
Qy 1000 PheValHisPheIleAlaAsp 1006
Db 3338 TTGTGCACTTTATCGCTGAC 3358
```

RESULT 2

```
US-10-104-047-314
; Sequence 314, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 314
; LENGTH: 3123
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-314

Alignment Scores:
Pred. No.: 1,17e-102 Length: 3123
Score: 110.00 Matches: 165
Percent Similarity: 98.80% Conservative: 0
Best Local Similarity: 98.80% Mismatches: 1
Query Match: 10.93% Indels: 2
DB: 17 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-104-047-314 (1-3123)
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 272 CCCCTGACCCCAAGCGCGCCGCCACCGCTTGGCAAGACGCCAGCGTAATGGAAGCCTTG 331
Qy 861 SerGlnProSerLysProAlaProProGlyLysLeuSerGlnLysArgProProLeuPro 880
Db 332 AGCCAGCGGAGCAAGCCTGCCCGCTGGGATCTCAGATCAGGCCCCCACCTCTTGCCC 391
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSerTh 900
Db 392 CCACAGCGCCAGCGCCCTCCCGCAGAGAAGCCTTGCGCGGGGCT-GACAAAGTCCAC 450
Qy 900 rProLeuThrAsnLysGlyGlnProProArgGlyProValAspLeuSerAlaThrGluAlaLe 920
Db 451 CCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCT 510
Qy 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 511 GGGTCTCTGTGCAATGCTATGGTCTTGCAAGCCCCCTGCAAGCCCCCTAGGAAGTCCGA 570
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 571 GGCACCAAGTTGAGGCTTAAGCGGTGAAGCGCTCTATAACTGTGTGGCTGACAAACC 630
Qy 960 oAspGluLeuThrPheSerGluGlyAspVallelleValAspGlyGluGluAspGlnG1 980
Db 631 CGATGAGCTACCTTCTCCAGGGGATGTGATCATCTGCGAGGAGGAGGAGGAGGAGGAG 690
Qy 980 uTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 691 GTGGTGGATTGGCCACATTGATGGAGATCTGTGGTTCGCAAGGCGCATTCCTCCGGTGTGCT 750
Qy 1000 eValHisPheIleAlaAsp 1006
Db 751 TGTGCACTTTATCGCTGAC 769

RESULT 3
US-09-764-875-37
; Sequence 37, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PUZ02
```



```
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 3436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-37

Alignment Scores:
Pred. No.: 1,28e-102 Length: 3436
Score: 110.00 Matches: 182
Percent Similarity: 98.91% Conservative: 0
Best Local Similarity: 98.91% Mismatches: 1
Query Match: 10.93% Indels: 2
DB: 11 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-764-875-37 (1-3436)

Qy 612 GlyAsnLeuAppLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThr 631
Db 1 GGGAACTGGATAAACAGACAGAGGAAAGGACAGACAGCCCTGCACCTACTGTGCTGACC 60
Qy 632 AspAsnAlaGluCysLeuLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsn 651
Db 61 GACAATGCCGAGTGCCTCAAGTTGCTCTCGGGGGAAGGCTTCATCGAGATAGCAAAAC 120
Qy 652 GluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGlu 671
Db 121 GAGTCAGAGAGACTCCGCTGGACATTGCCAAGCGCCTCAAGCAGACGACTGTGAGGAG 180
Qy 672 LeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrp 691
Db 181 CTGCTGACCCAAAGCCTTATCTGGAAGATTAAATCTCACGTTTCACTTCAATATGATGG 240
Qy 692 ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro 711
Db 241 CGACTACTCCAGAGAGACTGGATGAAAGTGAATGACGACATGGATGAGAAATTTGACGCCC 300
Qy 712 SerGlu-AsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLe 731
Db 301 AGTCC-CAACCGCGGGAAGACCGGCCCATCAGCTTCTACGAGCTGGGCTCCAACGACT 359
Qy 731 uGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnAr 751
Db 360 TCAGTCTAACCGCTGATCTTTGGCCAGAGATGCTGCAAAACCTTGCCAAAGGAGAAGCAGAG 419
Qy 751 gAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPr 771
Db 420 GCCTTTCATGCCAGCATCTTCAGATGAGACTTACGAGGCGCTTCTGAGTGGCAGGCC 479
Qy 771 oProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAs 791
Db 480 ACCTCCCGCCAGCCTGCGAGCCCCCAGCACCACGAGCGCCCCCGCTTCTCTCCAGGAA 539
Qy 791 nValGlyLys 794
Db 540 TGTGTGGCAAA 549

RESULT 4
US-09-764-875-358
; Sequence 358, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 358
; LENGTH: 3413
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-358

Alignment Scores:
Pred. No.: 1,53e-70 Length: 3413
Score: 79.00 Matches: 157
Percent Similarity: 97.52% Conservative: 0
Best Local Similarity: 97.52% Mismatches: 2
Query Match: 7.85% Indels: 4
DB: 11 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-764-875-358 (1-3413)

Qy 848 ProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAla 867
Db 572 CCACCCGTTGCCAAGACGCCCGCAGCGTAATGGAAGCCTTGAGCCAGCCGAGCAAGCTGCC 631
Qy 868 ProProGlyLysSerGlnIleArgProProProProProProProProProProProPro 887
Db 632 CCGCTCTGGGATCTCACAGATCAGGCCCGCCACCTCTGCCCGCCACAGCGCCGCGCTC 691
Qy 888 ProGlnLysLysProAlaPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyG1 907
Db 692 CCGCAGAAGAAGCCTGCGCGGGGGCT-GACAGTGCACCCCTGACCAACAAGGCCA 750
Qy 907 nProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMe 927
Db 751 ACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCTGGGTCTCTGTCCAATGCTRT 810
Qy 927 t-ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProL 947
Db 811 -GGTCTCTGCAGCCCGCTGCACCCATGCTAGGAAGTGCAGGCAACCAAGTTGAAGCCTA 869
Qy 947 ysArgValLysAlaLeuTyrAsnCyValAlaAspAsnProAspGluLeuThrPheSerG 967
Db 870 AGCGGGTGAAGCGCTCTATTAACGTGTGGCTGACAAACCCCGATGARTCACTTCTCCG 929
Qy 967 luGlyAspValIleValAspGlyGluGluAspGlnGluTrpIleGlyHisIleA 987
Db 930 AGGGGGATGTGATCATCTGTGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 989
Qy 987 spGlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 990 ATGAGAGATCTCTGTGTCGCAAGGCGCATTTCCCGGTGTCAATTTGTGCACCTTTATCGCTGAC 1048

RESULT 5
US-09-918-995-27010
; Sequence 27010, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27010
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(442)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27010
```

Alignment Scores:

Pred. No.: 2.53e-70 Length: 442
 Score: 78.00 Matches: 78
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.75% Indels: 0
 DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-918-995-27010 (1-442)

Qy 365 CysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGlnGlu 384
 Db TGCITTGACCTCAITTCATGACAGAACTTACCACCTTCAAGCTGAAGATGAACAGGAA 97
 Qy 385 CysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPhe 404
 Db TGTCAATATGGATGCTGTGCTGCAAAATAGCAAGAAGAGCTTTAAACAATGCAATT 157
 Qy 405 LysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIle 424
 Db AAGGGGATGACAACTACTGGAAGAAATAACATCGTCCAAAGAACTGACAAAGGAGATCATC 217
 Qy 425 SerGluValGlnArgMetThrClyAsnAspValCysCysAspCysGlyAlaPro 442
 Db TCAGAAGTGCAGAGATGACGGCAATGACGTCTGTGACTGTGGGGCGCCA 271

RESULT 6

US-09-815-343-1154
 ; Sequence 1154, Application US/09815343
 ; Patent No. US20010055596A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meagher, Madeleine
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: King, Gordon E.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.504
 ; CURRENT APPLICATION NUMBER: US/09/815,343
 ; CURRENT FILING DATE: 2001-03-22
 ; NUMBER OF SEQ ID NOS: 1556
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1154
 ; LENGTH: 563
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(563)
 ; OTHER INFORMATION: n = A,T,C or G

Alignment Scores:

Pred. No.: 3.53e-39 Length: 563
 Score: 48.00 Matches: 131
 Percent Similarity: 96.32% Conservative: 0
 Best Local Similarity: 96.32% Mismatches: 2
 Query Match: 4.77% Indels: 5
 DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-815-343-1154 (1-563)

Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
 Db AACATGCTATTAAGGGGATGACAACTACTGGAGAAAATAACATCGTCCAAAGAACTGACA 61
 Qy 421 LysGluIleIleSerGluValGln-ArgMetThrGlyAsnAspValCysCysAspCysG1 440
 Db AAGGAGATCATCTCAGAGTGCT-GAGGATGACGGCAATGACGTCTGTGACTGTGG 120
 Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
 Db GCGCCAGATCCTACATGGCTTTCACCAACCTGGGCATCTGACCTGATCGAGTGTTC 180

Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspV 480
 Db CGGAATCCACCGAGAGCTGGGGTTTCATTATTCAG-CATGCAGTCCCTGACCTTAGATG 239
 Qy 480 alLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500
 Db TACTGGGAACATCTGAGCTGCTGCGCCAGAATAATTGGGAATCGAGCTTTAATGAGA 299
 Qy 500 leMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer-AspMet 519
 Db TCATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCACCCAGGAGCCGACATG 359
 Qy 520 AsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 533
 Db AATGCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGG 401

RESULT 7

US-10-097-105-1154
 ; Sequence 1154, Application US/10097105
 ; Publication No. US20040037842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meagher, Madeleine Joy
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Xu, Jiangchun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.504C1
 ; CURRENT APPLICATION NUMBER: US/10/097,105
 ; CURRENT FILING DATE: 2002-03-13
 ; NUMBER OF SEQ ID NOS: 1562
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1154
 ; LENGTH: 563
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: 403-446, 464, 465, 472, 483, 485, 499, 539, 554
 ; OTHER INFORMATION: n = A,T,C or G

Alignment Scores:

Pred. No.: 3.53e-39 Length: 563
 Score: 48.00 Matches: 131
 Percent Similarity: 96.32% Conservative: 0
 Best Local Similarity: 96.32% Mismatches: 2
 Query Match: 4.77% Indels: 5
 DB: 18 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-097-105-1154 (1-563)

Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
 Db AACATGCTATTAAGGGGATGACAACTACTGGAGAAAATAACATCGTCCAAAGAACTGACA 61
 Qy 421 LysGluIleIleSerGluValGln-ArgMetThrGlyAsnAspValCysCysAspCysG1 440
 Db AAGGAGATCATCTCAGAGTGCT-GAGGATGACGGCAATGACGTCTGTGACTGTGG 120
 Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
 Db GCGCCAGATCCTACATGGCTTTCACCAACCTGGGCATCTGACCTGATCGAGTGTTC 180
 Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspV 480
 Db CGGAATCCACCGAGAGCTGGGGTTTCATTATTCAG-CATGCAGTCCCTGACCTTAGATG 239
 Qy 480 alLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500
 Db TACTGGGAACATCTGAGCTGCTGCGCCAGAATAATTGGGAATCGAGCTTTAATGAGA 299

```
Oy 500 leMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer-AspMet 519
Db 300 TCATGGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGCGACATG 359
Oy 520 AsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 533
Db 360 ATCTCAAGAAAGGACTATCATCCAGCCCAAGTACATCGAGAGG 401

RESULT 8
US-10-027-632-34151
; Sequence 34151, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34151
; LENGTH: 958
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34151

Alignment Scores:
Pred. No.: 1.29e-28 Length: 958
Score: 38.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.78% Indels: 0
DB: 13 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-027-632-34151 (1-958)
Oy 673 LeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArg 692
Db 41 CTGACCCCAAGCCTTATCTGGAGATTAAATCTCACGTTTCACTGATGATGAGAAATTCGAG 100

Oy 693 LeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
Db 101 CTACTCCACGAAGACCTGGATGAAAGTGTATGACGACATGATGAGAAATTCGAG 154

RESULT 9
US-10-027-632-34151
; Sequence 34151, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

```
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34151
; LENGTH: 958
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34151

Alignment Scores:
Pred. No.: 1.29e-28 Length: 958
Score: 38.00 Matches: 38
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.78% Indels: 0
DB: 17 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-027-632-34151 (1-958)
Oy 673 LeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArg 692
Db 41 CTGACCCCAAGCCTTATCTGGAGATTAAATCTCACGTTTCACTGATGATGAGAAATTCGAG 100

Oy 693 LeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
Db 101 CTACTCCACGAAGACCTGGATGAAAGTGTATGACGACATGATGAGAAATTCGAG 154

RESULT 10
US-10-276-774-404/c
; Sequence 404, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 404
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-404

Alignment Scores:
Pred. No.: 1.12e-20 Length: 409
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 18 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-276-774-404 (1-409)
Oy 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysProphe 146
```


; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22014
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-22014

Alignment Scores:
Pred. No.: 4.63e-10 Length: 60
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.89% Indels: 0
DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-908-975-22014 (1-60)

Qy 525 TyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHisAlaAspAsn 543
Db 3 TACATCAGCGCAAGTACATCGAGAGGAGATACGCAAGGAAGACGCGGATAAC 59

RESULT 14

US-09-815-343-1151
; Sequence 1151, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1151
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(326)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-1151

Alignment Scores:
Pred. No.: 2.58e-07 Length: 326
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-815-343-1151 (1-326)

Qy 411 GlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluVal 427
Db 32 GGAGAAATAAATACATCGTCCAAAGAACTGACAAAGGAGATCATCTCAGAAAGTG 82

RESULT 15

US-10-097-105-1151
; Sequence 1151, Application US/10097105
; Publication No. US20040037842A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; APPLICANT: Secrist, Heather
; APPLICANT: Harlocker, Susan L.

; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.504C1
; CURRENT APPLICATION NUMBER: US/10/097,105
; CURRENT FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 1562
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1151
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14, 29, 85, 105, 149, 150, 173, 208, 220, 223, 240, 252,
; LOCATION: 256, 286, 285, 299, 302, 304, 318, 322
; OTHER INFORMATION: n = A,T,C or G
US-10-097-105-1151

Alignment Scores:
Pred. No.: 2.58e-07 Length: 326
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 18 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-097-105-1151 (1-326)

Qy 411 GlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluVal 427
Db 32 GGAGAAATAAATACATCGTCCAAAGAACTGACAAAGGAGATCATCTCAGAAAGTG 82

Search completed: August 5, 2005, 04:01:52
Job time : 1466 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 17:15:00 ; Search time 393 Seconds
(without alignments)
4188.535 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDGPRKGAFVSVFHFAD 1006

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Word size: 1

Total number of hits satisfying chosen parameters: 2400006

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_b/US09914042/runat_04082005_090748_2684/app.query.fasta_1.1159
-DB-Issued_Patents=NA -QMT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=psto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914042 @CGN 1.1 80 @runat_04082005_090748_2684 -NCPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	29.5	4486	1	US-08-322-742-15
2	257	25.5	4328	1	US-08-322-742-12
3	68	6.8	2949	4	US-09-023-905A-8
4	68	6.8	5954	4	US-09-023-905A-6
5	32	3.2	2712	4	US-09-023-905A-11
6	32	3.2	4595	4	US-09-023-905A-9
7	30	3.0	3456	4	US-09-023-905A-5
8	30	3.0	4382	4	US-09-023-905A-3
9	30	3.0	5330	4	US-09-023-905A-1
10	16	1.6	213	4	US-09-513-999C-22310
11	12	1.2	4190	3	US-08-938-291A-2
12	12	1.2	4190	4	US-09-589-619-2

13	10	1.0	417	4	US-09-452-709A-140
c 14	10	1.0	697	4	US-09-452-709A-446
15	9	0.9	332	4	US-09-513-999C-384
16	9	0.9	514	4	US-09-854-133-660
17	9	0.9	601	4	US-09-949-016-150613
18	9	0.9	601	4	US-09-949-016-150614
19	9	0.9	601	4	US-09-949-016-150615
20	9	0.9	909	4	US-09-252-991A-12690
21	9	0.9	1131	3	US-09-383-586-7
22	9	0.9	1131	4	US-09-823-038A-7
23	9	0.9	1149	4	US-09-252-991A-12529
24	9	0.9	1295	1	US-08-433-854-3
25	9	0.9	1295	1	US-08-174-745A-3
26	9	0.9	1295	2	US-08-195-947-3
27	9	0.9	1295	2	US-08-433-885-3
28	9	0.9	1295	2	US-08-433-908B-3
29	9	0.9	1295	3	US-08-410-614-3
30	9	0.9	1640	4	US-09-573-080A-144
31	9	0.9	2119	4	US-09-620-312D-35
32	9	0.9	2366	4	US-09-270-767-13819
33	9	0.9	6990	4	US-09-620-312D-155
34	9	0.9	7158	4	US-09-949-016-14106
35	9	0.9	15378	3	US-08-785-420-1
36	9	0.9	37254	4	US-09-949-016-15973
c 37	9	0.9	38575	4	US-09-949-016-17304
38	9	0.9	38772	4	US-09-949-016-12729
39	9	0.9	38772	4	US-09-949-016-12729
40	9	0.9	77618	4	US-09-949-016-11768
41	9	0.9	94748	4	US-09-949-016-12648
c 42	9	0.9	94758	4	US-09-949-016-16741
c 43	9	0.9	108440	4	US-09-949-016-12065
c 44	9	0.9	108441	4	US-09-949-016-14090
45	9	0.9	113379	4	US-09-949-016-17561

ALIGNMENTS

RESULT 1

US-08-322-742-15
; Sequence 15, Application US/08322742
; Patent No. 5688641
; GENERAL INFORMATION:
; APPLICANT: Sager, Ruth
; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,742
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/938,823
; FILING DATE: September 1, 1992
; APPLICATION NUMBER: 07/844,296
; FILING DATE: February 28, 1992
; APPLICATION NUMBER: 07/552,216
; FILING DATE: February 28, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/048003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070

Sequence 140, App	Sequence 140, App
Sequence 446, App	Sequence 446, App
Sequence 384, App	Sequence 384, App
Sequence 660, App	Sequence 660, App
Sequence 150613,	Sequence 150613,
Sequence 150614,	Sequence 150614,
Sequence 150615,	Sequence 150615,
Sequence 12690, A	Sequence 12690, A
Sequence 7, Appli	Sequence 7, Appli
Sequence 12529, A	Sequence 12529, A
Sequence 3, Appli	Sequence 3, Appli
Sequence 3, Appli	Sequence 3, Appli
Sequence 3, Appli	Sequence 3, Appli
Sequence 3, Appli	Sequence 3, Appli
Sequence 144, App	Sequence 144, App
Sequence 35, Appli	Sequence 35, Appli
Sequence 13819, A	Sequence 13819, A
Sequence 155, App	Sequence 155, App
Sequence 14106, A	Sequence 14106, A
Sequence 1, Appli	Sequence 1, Appli
Sequence 15973, A	Sequence 15973, A
Sequence 17304, A	Sequence 17304, A
Sequence 12382, A	Sequence 12382, A
Sequence 12729, A	Sequence 12729, A
Sequence 11768, A	Sequence 11768, A
Sequence 12648, A	Sequence 12648, A
Sequence 16741, A	Sequence 16741, A
Sequence 12065, A	Sequence 12065, A
Sequence 14090, A	Sequence 14090, A
Sequence 17561, A	Sequence 17561, A

; TELEFAX: (617) 542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4486
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 US-08-322-742-15

Alignment Scores:

Pred. No.: 6,46e-289 Length: 4486
 Score: 297.00 Matches: 701
 Percent Similarity: 98.59% Conservative: 0
 Best Local Similarity: 98.59% Mismatches: 4
 Query Match: 29.52% Indels: 10
 DB: 1 Gaps: 0

US-09-914-042-1 (1-1006) x US-08-322-742-15 (1-4486)

Qy 302 LysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLys 321
 Db 1 AAGGAACATGGACCGCGAAGCGGAGCTCTACAAAGAGAGTGACGGGATCCGAAAA 60
 Qy 322 ValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341
 Db 61 GTGTGGCAGAAAGGAAATGTTTCAGTTAAATATGTTTCTGACCATATCCCATGGTACC 120
 Qy 342 AlaAsnArg-ProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProG1 361
 Db 121 GCTAACCGGGCTCTCTGCAAGCTCAACTGCTTAACCTGCCAGGTGAAGACCAACCTGA 180
 Qy 361 uGluLysLysCysPheAspLeuLysSerHisAspArgThrTyrHisPheGlnAlaGluAs 381
 Db 181 GGAGAAGAAGTGCTTTTGACCTTATTTTCATCATGACAGAACTTACCACCTTTCAAGCTGAAGA 240
 Qy 381 pGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluAlaLeuAs 401
 Db 241 TGAACAGGAATGTCAAAATATGATGTCTGTGCTGCAAAATAGCAAGAAGAGCTTTAAA 300
 Qy 401 nAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys 421
 Db 301 CAATGCATTTAAGGGGGATGACAACTATGGAGAAATAACATCGTCCAGAACTGACAAA 360
 Qy 421 sGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspGlyAl 441
 Db 361 GGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGAGTGTGGGC 420
 Qy 441 aProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerG1 461
 Db 421 GCCAGATCTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATCGAGTGTTCGG 480
 Qy 461 VileHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspValL 481
 Db 481 AATCACCAGAGAGCTGGGGTTCATTATTCCAG-GATGCAGTCCCTGACCTAGATGTAC 539
 Qy 481 euGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleM 501
 Db 540 TGGGAACATCTCAGCTGTGCTCGCGCAAGAAATATTGGGAATTCAGGCTTTAATGAGATCA 599
 Qy 501 etGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnA 521
 Db 600 TGGAAATGTTCCTACAGCTGAGGACTCAGTCAAAACCCCAACCCAGCGAGCATGAATG 659
 Qy 521 IaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisA 541
 Db 660 CAAGAAAGGACTATCATCAGCCCAAGTACATCAGAGGAGATACCGAAGAGAGACAG 719
 Qy 541 IaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheG 561
 Db 720 CGGATAACCGCGGAGAGCTTTCAGTCTTTTGGAGCGCTCAAAAACGAGAGATATTTTGT 779
 Qy 561 LysLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnG 581

Db 780 GATTGCTCCAAGCTTATGCTGATGGTGTGGATCTTACGGAATAATCCACTGCGCAACG 839
 Qy 581 LysGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerL 601
 Db 840 GACATGAGCCGGATGAAACCGCCCTCCACCTTGCACTCGAGTCAGATCGGTGGATCGAACCTCTC 899
 Qy 601 euHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysG 621
 Db 900 TTCACATTTGTAGACTTTTATGTTTTCAGAACAGTGGGAACCTGGATATAACAGACAGGGAAG 959
 Qy 621 LysThrAlaLeuHisTyrCysValLeuThrAspAsnAlaGluCysLeuLysLeuLeuL 641
 Db 960 GCAGCAGCCCTCGACTACTGCTGCTGACCGCAATGCGAGTGCCTCAAGTTGCTTCC 1019
 Qy 641 euArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleA 661
 Db 1020 TCGGGGGAAGGCTCTCATCGAGATAGCAATATGATCAGGAGAGACTCCGCTGGACATTG 1079
 Qy 661 LaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgP 681
 Db 1080 CCAAGCCCTCAAGCAGCAGCACTGTGAGGAGCTGTGACCCAGCCCTTATCTGGAAGAT 1139
 Qy 681 heAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGluS 701
 Db 1140 TTAATTTCTCAGTTCACTGTTGAATATGAATGGCCACTACTCCACGAGACCTGGATGAA 1199
 Qy 701 erAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArgPro 720
 Db 1200 GTGATGACGACATGGATGAGAAATTTGACGCCAGTTC-CAACCGCGGGAAGACCGGCC 1258
 Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
 Db 1259 ATCAGCTTCTACAGCTGGGCTCCCAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 1318
 Qy 741 AspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPheMetProSerIleLeuGlnAs 760
 Db 1319 GATGCTGCAAACTTTGCCNAGGA-CAAGCAGAGGGCTTTTCATGCCAGCATCTTGCAGAA 1377
 Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAProSe 780
 Db 1378 TGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCCACTCCCGCCAGCTGAGCCGCCAG 1437
 Qy 780 rThrThrSerAla-ProProLeuProProArgAsnValGlyValGlnThrAlaSerS 800
 Db 1438 CACCACAGCGCCCGCCCGCTTCTCCACGGAATTTGGCAAAAGTTTCAGACAGCTCTCT 1497
 Qy 800 erAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS 820
 Db 1498 CTGCTAACACCTGTGGNAGACAACTCTGTAGTGTGGACGGTGGAGCGCGGAGCGCAT 1557
 Qy 820 erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA 840
 Db 1558 CTTCGTTCAGATCCGCGAGCTGTCCATCCACGCTGCGCCCTCTTCGCGTGACATCTACCA 1617
 Qy 840 snProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaL 860
 Db 1618 ATCCCTTACCCCGCCCGCCCGCTTCCAGGAGTGGCAAGACCGCCAGCGCTTATGGAAGCT 1677
 Qy 860 euSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuP 880
 Db 1678 TGAGCCAGCGAGCAAGCTGCCCCGCTGGGATCTCACAGATCAGGCCCGCCACCTCTGC 1737
 Qy 880 roProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer 899
 Db 1738 CCCCACAGCGCCCGCCCGCTCCCGCAGAGAAGAGCTTCGCGCAGGGGCT-GACAAGTCC 1796
 Qy 900 ThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAla 919
 Db 1797 ACCCCACTGACCAACAAAGGCCAACCCGAGAGAGACTGTGGATCTCTCTGCAACCGAAGCT 1856
 Qy 920 LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer 939

Db 1857 CTGGGTCCTCTGTCCTCAATGCTATGTCCTGAGAGCCCTGACCCATGCTAGGAGTCG 1916
 Qy 940 GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsn 959
 Db 1917 CAGGCAACCAAGTTCAAGCCTAAGCGGTGAAAGCGCTCTATAACTGTGTGGTGACAAC 1976
 Qy 960 ProAspGluLeuThrPheSerGluClyAspValIleIleValAspGlyGluGluAspGln 979
 Db 1977 CCCGATGAGCTCACCTCTCCGAGGGGATGTGATCATCGTGGACGGGAGGAGACCAG 2036
 Qy 980 GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer 999
 Db 2037 GAGTGGTGGATGGCCACATGATGGAGATCTGTGTCGCAAGGCCCATTTCCGGTGTC 2096
 Qy 1000 PheValHisPheIleAlaAsp 1006
 Db 2097 TTTGTGCATTTATCGCTGAC 2117

RESULT 2

US-08-322-742-12
 ; Sequence 12, Application US/08322742
 ; Patent No. 5688641
 ; GENERAL INFORMATION:
 ; APPLICANT: Sager, Ruth
 ; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM PS/2 Model 50Z or 55SX
 ; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 ; SOFTWARE: WordPerfect (Version 5.1)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/322,742
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/938,823
 ; FILING DATE: September 1, 1992
 ; APPLICATION NUMBER: 07/844,296
 ; FILING DATE: February 28, 1992
 ; APPLICATION NUMBER: 07/552,216
 ; FILING DATE: February 28, 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fraser, Janis K.
 ; REGISTRATION NUMBER: 34,819
 ; REFERENCE/DOCKET NUMBER: 00530/048003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 542-5070
 ; TELEFAX: (617) 542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4328
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-322-742-12

Alignment Scores:
 Pred. No.: 1,35e-248 Length: 4328
 Score: 257.00 Matches: 648
 Percent Similarity: 98.63% Conservative: 0
 Best Local Similarity: 98.63% Mismatches: 4
 Query Match: 25.55% Indels: 9
 DB: 1 Gaps: 0

US-09-914-042-1 (1-1006) x US-08-322-742-12 (1-4328)

Qy 355 GlnValLysThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr 374
 Db 3 CAGGTGAAGACCAACCCCTGAGGAGAAGTGTCTTGACCTTTATTCATCATGACAGAACT 62
 Qy 375 TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsn 394
 Db 63 TACCACCTTTCAAGCTGAAGATGAACAGGAATGTCAAAATATGGATGTCTGTCTGCAAAAT 122
 Qy 395 SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn 414
 Db 123 AGCAAGAGAGAGCTTTAAACAATGCATTTAAGGGGGATGACATATCTGGAGAAAATTAAC 182
 Qy 415 IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAsp 434
 Db 183 ATCGTCCAAGAACTGACAAAGAGATCATCTCAGAAGTGCAGAGGATGACGGCAATGAC 242
 Qy 435 ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu 454
 Db 243 GTCTGTGTGACTGTGGGGGCCAGATCTCATATGGCTTTCCACCAACCTGGGCATCTCG 302
 Qy 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetG1 474
 Db 303 ACCTGCATCGAGTGTTCGGGAATCCACGAGAGCTGGGGGTTCATTATTCAG-CATGCA 361
 Qy 474 nSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAs 494
 Db 362 GTCCCTGACCTTAGATGTACTGGGAACATCTGAGCTGCTGCTCGCCAAGATATTTGGGA 421
 Qy 494 nAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAs 514
 Db 422 TGCAGGCTTTAATGAGATCATGGAATGTGCTACCAAGTGCAGCTGAGGACTCAGTCAACCCAA 481
 Qy 514 nProGlySerAspMetAsnAlaAaLysLysAspTyrIleThrAlaLysTyrIleGluArgAr 534
 Db 482 CCCAGGCGACGACATGATGCAAGAAGAGACTACATCACGCCAAGTACATCGAGAGAG 541
 Qy 534 gTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVa 554
 Db 542 ATACGCAAGGAAGAAGCAGCGGATACCGCGCAAGCTTCACAGCTCTTTGCGAGGCCGT 601
 Qy 554 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrG1 574
 Db 602 CAAAACGAGAGATATTTTGGATTGCTCCAAAGCTTATGCTGATGGTGTGGATCTTTACGGA 661
 Qy 574 uLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr 594
 Db 662 AAAATCCCACTGGCCCAACGACATGAGCCGGATGAACGCGCCCTCCACCTTTGCAAGTCAG 721
 Qy 594 gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLe 614
 Db 722 ATCCGTGGATCGAACCTCTCTTCACATTTGTAGACTTTTGTAGTTCAGAACAGTGGGAACCT 781
 Qy 614 uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAl 634
 Db 782 GGATAAACAAGCAGAGGAAAGGACACAGCCCTGCACCTACTGCTGCTGAGCCGACAAATGC 841
 Qy 634 aGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerG1 654
 Db 842 CGAGTGCCTCAAGTTGCTCTTCGCGGGGAAGCCCTCCATCGAGATAGCAATAGTCTCAGG 901
 Qy 654 yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuTh 674
 Db 902 AGAGACTCCGCTGGACATTTGCCAAGCGCTCAAGCAGGACCTGTGAGGAGCTGCTGAC 961
 Qy 674 rGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLe 694
 Db 962 CCAAGCCTTATCTGGAAGATTTAATTTCTCAGTTTCAAGTTCAGATATGAATGGCCGACTACT 1021
 Qy 694 uHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlu-A 714
 Db 1022 CCACGAAGACCTGTGATGAAGTGTATGACGATGATGGATGAGAAATTTGCAGCCCGCTCC-CA 1080

```
Qy 714 snArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerA 734
Db 1081 ACCGGCGGAGAGACCGGCCCATCAGCTTCTACAGCTGGCTCCACAGCTTTCAGTCTA 1140
Qy 734 snAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPhe 753
Db 1141 ACGTGATATCTTTGGCCAGAGATGCTGCAACCTTGCCCAAGGA-CAAGCAGAGGGCTTTC 1199
Qy 754 MetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProPro 773
Db 1200 ATGCCACAGCTTTGCAAGATGAGACTTACGAGGCCCTCTGAGTGGCAGCCACCTCCC 1259
Qy 774 AlaGlnProAlaAlaProSerThrThrSerAla-ProProLeuProProArgAsnValG1 793
Db 1260 GCCAGCCTTGAGCCCGCCAGCACCCAGCCGCCCCCGCTTCTCCACGGAATGTTGG 1319
Qy 793 YLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAs 813
Db 1320 CAAAGTTTCAGACAGCCTCTCTGCTAAACACCTGTGGAAGACAAACTCTGTAAGTGTGA 1379
Qy 813 pGlyGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProPr 833
Db 1380 CGGTGGAAGCGGCGGCGATCTTCGTAGATCCGCGCAGCTGTCCATCCACGCTGCCGCC 1439
Qy 833 oLeuArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysTh 853
Db 1440 TCTTCGCGTGACATCTACCAATCCCTGACCCCGCCGCGCCCGCCCGCTTGCCAGAC 1499
Qy 853 rProSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLysSerG1 873
Db 1500 GCCAGCGCTAATGGAAGCTTTGAGCCAGCGGAGCAAGCTGCCCGCTGGGATCTACA 1559
Qy 873 nileArgProProProProLeuProProProProSerArgLeuProGlnLysLysProAl 893
Db 1560 GATCAGGCCCCCACCCTCTGCCCCCAGCCGCGCCAGCCGCTCCCGCAGAGAAGCCCTGC 1619
Qy 893 apro-GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValA 913
Db 1620 GCCAGGGGCT-GACAAAGTCCACCCCACTGACCAACAAAGGCCAACCCGAGAGACTGTGG 1678
Qy 913 spLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProA 933
Db 1679 ATCTCTCTGCAAGGGAAGCTCTGGGCTCTCTGTCTCAATGCTATGCTCTCGAGCCCTCG 1738
Qy 933 laProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuT 953
Db 1739 CACCCATCCCTAGGAAGTCGAGGCAACCAAGTTGAAGCCTTAAGCGGTGAAAGCGCTCT 1798
Qy 953 YrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleV 973
Db 1799 ATAAGTGTGGTGACAAACCCGATGAGCTCACCTTCTCGAGGGGATGTGATCATCG 1858
Qy 973 alAspGlyGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgL 993
Db 1859 TGGACGGGGAGGAGGACCAAGGAGTGGTGGATTGGCCACATTGATGGAGATCTCTGTGCA 1918
Qy 993 ysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 1919 AAGCGCATTCCTCCGCTGTCAATTTGTGCACCTTTATCGCTGAC 1959
```

RESULT 3

```
US-09-023-905A-8
; Sequence 8, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
```

```
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 2949
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-8

Alignment Scores: 3.53e-58 Length: 2949
Pred. No.: 68.00 Matches: 68
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 6.76% Indels: 0
Query Match: 4 Gaps: 0
DB:

US-09-914-042-1 (1-1006) x US-09-023-905A-8 (1-2949)

Qy 171 HisGlyMetIleArgThrGluIleSerGlyValaGluIleAlaGluMetGluLysGlu 190
Db 511 CACGGAATGATCCGACGGAGATCAGCGGAGCAGATAGCAGAAAGAGATGGAAGAGAG 570
Qy 191 ArgArgPhePheGlnLeuGlnMetCysGluTyrLysLeuLysValAsnGluIleLysIle 210
Db 571 CGGCGTTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAGATCAATGAATCAAGATC 630
Qy 211 LysLysGlyValaAspLeuLeuGlnAsnLeuLysIleLysTyrPheHisAlaGlnCysAsnPhe 230
Db 631 AAAAAAGGTGTGACCTGTCTCCAGAACTCATCAATACTTCCACGACAGTGCACACTTC 690
Qy 231 PheGlnAspGlyLeuLysAlaVal 238
Db 691 TTTCAGGATGGTCTCAAAAGCGGTG 714

RESULT 4
US-09-023-905A-6
; Sequence 6, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 5954
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (433) ... (3378)
US-09-023-905A-6

Alignment Scores: 7.07e-58 Length: 5954
Pred. No.: 68.00 Matches: 68
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00%
```

```
Query Match: 6.76% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-6 (1-5954)

Qy 171 HisGlyMetIleAqThrGluLeuSerGlyAlaGluIleAlaGluGluMetGluIleGlu 190
Db 943 CACGGAATGATCCGGACCGAGATACGCGAGCAGATAGCAGAAAGATGAAAAAGAG 1002
Qy 191 ArgArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIle 210
Db 1003 CGGGCTTTCTCCAGCTTCAGATGTGTGAGTACCTCTCAAAGTCAATGAATCAAGATC 1062
Qy 211 LysLysGlyValAspLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhe 230
Db 1063 AAAAAAGGTGCGACCTCTCAGAAATCTCATCAATAATCTTCCACGCACAGTGCACACTTC 1122
Qy 231 PheGlnAspGlyLeuLysAlaVal 238
Db 1123 TTTCAGGATGGTCTCAAAGCGGTG 1146

RESULT 5
US-09-023-905A-11
; Sequence 11, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 2712
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-11
Alignment Scores:
Pred. No.: 6.51e-22 Length: 2712
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.18% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-11 (1-2712)

Qy 125 PheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLys 144
Db 373 TTTCACATAGACAGTCTGCTTAAGGAGACCTCAAAGGAGTCAAAGGGGATTTGAAAAAG 432
Qy 145 ProPheAspLysAlaTrpLysAspTyrGluThrLys 156
Db 433 CCATTTGATAAGCATGGAAGGATTTATGAACCAA 468

RESULT 6
US-09-023-905A-9
; Sequence 9, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 4595
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-9
Alignment Scores:
Pred. No.: 1.1e-21 Length: 4595
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.18% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-9 (1-4595)

Qy 125 PheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLys 144
Db 672 TTTCACATAGACAGTCTGCTTAAGGAGACCTCAAAGGAGTCAAAGGGGATTTGAAAAAG 731
Qy 145 ProPheAspLysAlaTrpLysAspTyrGluThrLys 156
Db 732 CCATTTGATAAGCATGGAAGGATTTATGAACCAA 767

RESULT 7
US-09-023-905A-5
; Sequence 5, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3456
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-5
Alignment Scores:
Pred. No.: 8.6e-20 Length: 3456
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 4 Gaps: 0
```

US-09-914-042-1 (1-1006) x US-09-023-905A-5 (1-3456)

Qy 127 LeuAspSerLeuLeuGlyAspLeuLysGlyVallysGlyAspLeuLysProphe 146
Db 439 CTGGACTCTCTGCTCAAGGAGATCTAAAGGAGTGAAGGGGACCTTTAAAAAGCCTTTC 498
Qy 147 AspLysAlaTrpLysAspTyrGluThrLys 156
Db 499 GACAAGGCTGGAAGACTATGAACCAAG 528

RESULT 8

US-09-023-905A-3
; Sequence 3, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 4382
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (351)...(3803)
US-09-023-905A-3

Alignment Scores:
Pred. No.: 1.09e-19 Length: 4382
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-3 (1-4382)

Qy 127 LeuAspSerLeuLeuGlyAspLeuLysGlyVallysGlyAspLeuLysProphe 146
Db 789 CTGGACTCTCTGCTCAAGGAGATCTAAAGGAGTGAAGGGGACCTTTAAAAAGCCTTTC 848
Qy 147 AspLysAlaTrpLysAspTyrGluThrLys 156
Db 849 GACAAGGCTGGAAGACTATGAACCAAG 878

RESULT 9

US-09-023-905A-1
; Sequence 1, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5330
; TYPE: DNA
; ORGANISM: Bos sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (209)...(3595)
US-09-023-905A-1

Alignment Scores:
Pred. No.: 1.32e-19 Length: 5330
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-023-905A-1 (1-5330)

Qy 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyVallysGlyAspLeuLysProphe 146
Db 647 TTGGATTCTCTTGTGAAGGAGACCTGAAGGGAGCTCAAGGCGATCTCAAGAAACCATTT 706
Qy 147 AspLysAlaTrpLysAspTyrGluThrLys 156
Db 707 GACAAGCTTTGGAAGATTATGAGACGAAG 736

RESULT 10

US-09-513-999C-22310
; Sequence 22310, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 22310
; LENGTH: 213
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 33
; OTHER INFORMATION: v-a or c or g
US-09-513-999C-22310

Alignment Scores:
Pred. No.: 7.18e-07 Length: 213
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.59% Indels: 0
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-513-999C-22310 (1-213)

Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGlu 465
Db 36 AACTTGGGTATTTTGACCTGTATAGAAATGTTCTGGCATCCATAGGAA 83

RESULT 11

```

; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; storage
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/589,619
; FILING DATE: 07-Jun-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,291
; FILING DATE: September 26, 1997
; APPLICATION NUMBER: 60/027,337
; FILING DATE: October 11, 1996
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 228/172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4190 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
;
; US-09-589-619-2
;
; Alignment Scores:
;
; Pred. No.: 0.147 Length: 4190
; Score: 12.00 Matches: 12
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 1.13% Indels: 0
; DB: 4 Gaps: 0
;
; US-09-914-042-1 (1-1006) x US-09-589-619-2 (1-4190)
;
; QY 772 ProProAlaGInProAlaAlaProSerThrThrSer 783
;
; DB 2607 CCCCCGGCCAGCAGCGCGCCCCCAGCACCACCACT 2642
;
; RESULT 13
; US-09-492-709A-140
; Sequence 140, Application US/09492709A
; Patent No. 6720139
;
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohleen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
;
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0

```

; SEQ ID NO 140
; LENGTH: 417
; TYPE: DNA
; ORGANISM: E. Coli
US-09-492-709A-140

Alignment Scores:

Pred. No.:	1.57	Length:	417
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.99%	Indels:	0
DB:	4	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-492-709A-140 (1-417)

Qy 71 AenGluGluGlnTyrThrGlnAlaLeuGlu 80
|||||
Db 76 AACGAGGAACAATACACCCAGCGCTGGAA 105
|||||

RESULT 14

US-09-492-709A-446/c
; Sequence 446, Application US/09492709A
; Patent No. 6720139

GENERAL INFORMATION:

; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ESCHERICHIA COLI

; CURRENT APPLICATION NUMBER: US/09/492,709A

; CURRENT FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 446

; LENGTH: 697

; TYPE: DNA

; ORGANISM: Escherichia coli

US-09-492-709A-446

Alignment Scores:

Pred. No.:	2.6	Length:	697
Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.99%	Indels:	0
DB:	4	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-492-709A-446 (1-697)

Qy 71 AenGluGluGlnTyrThrGlnAlaLeuGlu 80
|||||
Db 241 AACGAGGAACAATACACCCAGCGCTGGAA 212
|||||

RESULT 15

US-09-513-999C-384

; Sequence 384, Application US/09513999C

; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 384
; LENGTH: 332

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 89..331

; FEATURE:

; NAME/KEY: sig_peptide

; LOCATION: 89..208

; OTHER INFORMATION: score 4.4

; OTHER INFORMATION: seq LNLGLVLMCIECSG/IH

US-09-513-999C-384

Alignment Scores:

Pred. No.:	12.7	Length:	332
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.89%	Indels:	0
DB:	4	Gaps:	0

US-09-914-042-1 (1-1006) x US-09-513-999C-384 (1-332)

Qy 456 CysleGluCysSerGlyIleHisArg 464
|||||
Db 191 TGTATTGAATGTTTCAGGAATCCACGC 217
|||||

Search completed: August 5, 2005, 00:25:15

Job time : 435 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 13:57:13 ; Search time 1228 Seconds
(without alignments)

4849.560 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHDYK.....DCGPKRGAPVSVFHFAD 1006

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8776198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters: -MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO spool_h/US09914042/runat_04082005_090746_2651/app.query.fasta_1.1159
-DB=N Geneseq -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914042 @CGN 1.1 615 @runat_04082005_090746_2651 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq 16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	257	25.5	4328	2	AQ28758 Partial s
2	110	10.9	3123	10	ADB62160 Human cDN
3	110	10.9	3436	4	ABK43447 DNA encod
4	110	10.9	3436	12	ADI53834 cDNA enco
5	79	7.9	3413	4	ABK43768 DNA encod

6	79	7.9	3413	12	ADI54155	Adi54155 cDNA enco
7	78	7.8	442	9	ACH39798	Ach39798 Human foe
8	68	6.8	5954	2	AAV59105	AAV59105 Zebrafish
9	48	4.8	563	4	AA58478	AA58478 CDNA #115
10	32	3.2	4595	2	AAV59106	AAV59106 Zebrafish
11	30	3.0	409	4	ABA08628	ABA08628 Human GTP
12	30	3.0	3775	10	ADD71214	Add71214 Human int
13	30	3.0	3946	10	ADD71207	Add71207 Human int
14	30	3.0	3974	12	ADJ51008	Adj51008 Human cDN
15	30	3.0	4382	2	AAV59104	AAV59104 Zebrafish
16	30	3.0	5330	2	AAV59103	AAV59103 Bovine di
17	30	3.0	5947	13	ADS34277	Ads34277 POSH prot
18	24	2.4	544	13	ADQ79413	Adq79413 Novel can
19	22	2.2	4870	13	ADS34272	Ads34272 POSH prot
20	22	2.2	4997	13	ADS10088	Ads10088 Human the
21	22	2.2	5475	13	ADS34273	Ads34273 POSH prot
22	21	2.1	345	4	AAI84720	Aai84720 Human pol
23	19	1.9	60	6	ABN49266	Abn49266 Human spl
24	17	1.7	326	4	AA58475	AA58475 CDNA #115
25	17	1.7	747	6	ABK71808	Abk71808 Human dit
26	16	1.6	213	3	AAAC18235	Aac18235 Human sec
27	16	1.6	469	4	ABA08512	AbA08512 Human dif
28	14	1.4	1038	4	ABA08863	AbA08863 Human sec
29	14	1.4	5033	4	AAI59224	Aai59224 Human pol
30	13	1.3	167	4	AAI46525	Aai46525 Probe #15
31	13	1.3	167	4	AKA40054	Aka40054 Human bon
32	13	1.3	167	4	ABS40054	AbA40054 Human liv
33	13	1.3	167	5	AAI06958	Aai06958 Probe #59
34	13	1.3	323	6	ABQ97783	Abq97783 Mouse ES
35	13	1.3	446	4	AAI33385	Aai33385 Probe #20
36	13	1.3	446	4	AAK27473	Aak27473 Human bon
37	13	1.3	446	5	AAI01973	Aai01973 Probe #19
38	12	1.2	296	6	ABL93107	AbI93107 Rat metas
39	12	1.2	296	12	ADN07897	Adn07897 Human mam
40	12	1.2	2503	3	AAAC69461	Aac69461 Human sec
41	12	1.2	3580	12	ADQ86427	Adq86427 Human tum
42	12	1.2	3580	13	ADQ85284	Adq85284 Human tum
43	12	1.2	3580	13	ADQ83337	Adq83337 Human tum
44	12	1.2	3710	13	ADR06841	Adr06841 Full leng
45	12	1.2	3710	13	ADR06841	ADR06841 Full leng

ALIGNMENTS

RESULT 1
AAQ28758
ID AAQ28758 standard; DNA; 4328 BP.
XX AC AAQ28758;
XX AC
XX 25-MAR-2003 (revised)
DT 25-FEB-1993 (first entry)
XX
XX Partial sequence of tumour suppressor gene U10.
DE
KW Can19; tumour suppressor gene; cancer; therapy; ss.
XX Homo sapiens.
XX
XX WO9215602-A1.
XX
XX 17-SEP-1992.
XX
XX 28-FEB-1992; 92WO-US001624.
XX
XX 28-FEB-1991; 91US-00662216.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX
XX Sager R;
XX WPI; 1992-331663/40.
XX

PT Diagnosis and treatment of cancer - using candidate tumour suppressor
 PT genes or the corresp. antibodies.
 XX
 PS Claim 30; Page 38-43; 54pp; English.

XX An adaptation of the subtractive hybridization technique was used which
 CC utilizes a biotinylation-based subtraction procedure instead of
 CC hydroxyapatite as previously used. In this procedure, a single strand
 CC phagemid cDNA library from normal cell polyA+ mRNA is hybridized with
 CC excess biotinylated tumor polyA+ mRNA, and the resulting double stranded
 CC sequences are removed by binding to streptavidin. The remaining single-
 CC stranded phagemid cDNAs are converted to double-stranded form and used to
 CC transform bacterial host cells. The resulting subtracted cDNA library is
 CC differentially screened with total cDNA from normal and tumor cells. This
 CC method produced some 20 additional cloned cDNAs. Also found by this
 CC method were several genes which, on the basis of the partial DNA
 CC sequences appear to be novel sequences not previously entered into
 CC GENBANK. The portion of the cDNAs so sequenced represents part of the
 CC coding region and/or part of the 3' untranslated region of each cDNA (see
 CC Q28749-58). (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-
 CC MAR-2003 to correct PA field.)

SQ Sequence 4328 BP; 1236 A; 970 C; 912 G; 1210 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.74e-243	Length:	4328
Score:	257.00	Matches:	648
Percent Similarity:	98.63%	Conservative:	0
Best Local Similarity:	98.63%	Mismatches:	4
Query Match:	25.55%	Indels:	9
DB:	2	Gaps:	0

US-09-914-042-1 (1-1006) x AAQ28758 (1-4328)

Qy	355	GlnValLysThrAsnProGluLysLysCysPheAspLeuLeuSerHisAspArgThr	374
Db	3	CAGGTGAAGACCAACCCCTGAGGAGAAAGTCTTTGACCTTATTCATGACAGAACT	62
Qy	375	TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTropMetSerValLeuGlnAsn	394
Db	63	TACCACCTTTCAGCTGAAGATGAAACAGGAATGTCAAAATATGATGCTGTGCTGCAAAAT	122
Qy	395	SerLysGluGluAlaLeuAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn	414
Db	123	AGCAAGAAAGAAAGCTTTAAACATCATTTAAAGGGGATGACATCTGGAGAAATTAAC	182
Qy	415	IleValGlnGluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAsp	434
Db	183	ATCGTCCAAGAACTGACAAAGGAGATCATCTCAGAAAGTCAGAGGATGACGGGCAATGAC	242
Qy	435	ValCysCysAspCysGlyValaProAspProThrTropLeuSerThrAsnLeuGlyIleLeu	454
Db	243	GTCTGCTGTGACTGTGGGGCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTGT	302
Qy	455	ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetG1	474
Db	303	ACCTGCATCGAGTGTTCGGGAATCCACCGAGAGCTGGGGGTTCATTATTCAG-CATGCA	361
Qy	474	nSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAs	494
Db	362	GTCCCTGACCTTAGATGTACTGGGAACATCTGAGCTGTGCTGCGCAAGATATTTGGAA	421
Qy	494	nAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAs	514
Db	422	TGCAGGCTTTAATGAGATCATGGAATGTGGCTTACCAGCTGAGGACTCAGTCAACCCAA	481
Qy	514	nProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgAr	534
Db	482	CCCGCAGCGCATGATGCAAGAAAGGACTTACATCAGCCAGTACATCGAGAGAG	541
Qy	534	gTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVa	554
Db	542	ATACGCAAGGAAGACGCGGATAACGCGCGGAAGCTTTCACAGTCTTTTGCAGGCGGT	601

Qy	554	lLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValaPheLeuThrG1	574
Db	602	CAAAACGAGAGATATTTTGGATTGCTCCAAAGCTTATGCTGATGTTGGTATCTTACGGA	661
Qy	574	uLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr	594
Db	662	AAAAATCCCACTGGCCCAACGACATGATGCGGATGAACGGCCCTCCACCTTGCAGTCAG	721
Qy	594	gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLe	614
Db	722	ATCCGTGGATCGAAACCTCTCTTCATATTGTAGACTTTTGTAGTTTTCAGAACAGTGGGAACCT	781
Qy	614	uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAl	634
Db	782	GGATAAACAGACAGGGAAGGACACAGCCCTGCATCTACTGCTGCTGACCCGCAATGC	841
Qy	634	gGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerG1	654
Db	842	CGAGTGCCTCAAGTTGCTCTGCGGGGGAAGGCCCTCCATCGAGATAGCAAAATGAGTCAGG	901
Qy	654	yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuTh	674
Db	902	AGAGACTCCGCTGGACATTGCCAAGCGCCTCAAGCACGACGACTGTGAGGAGCTGTGCAC	961
Qy	674	xGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgLeuLe	694
Db	962	CCAAAGCCTTATCTCGAAGATTTAATTTCTACGTTTCACTGTTGAATATGAAATGGCCACTACT	1021
Qy	694	uHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlu-A	714
Db	1022	CCACGAAGACCTGGATGAAAGTATGACGATGATGAGAAATTCAGCCAGCTCC-CA	1080
Qy	714	snArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerA	734
Db	1081	ACCGCGGGAAGACCGGCCCATCATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCAGTCTA	1140
Qy	734	snAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPhe	753
Db	1141	ACGCTGTATCTTTGGCCAGAGATGCTGCTGCAAACTTTGCCAAGGA-CAAGCAGAGAGGCTTTC	1199
Qy	754	MetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProPro	773
Db	1200	ATGCCAGCATCTTTGCAGATGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCACCTCCC	1259
Qy	774	AlaGlnProAlaAlaProSerThrThrSerAla-ProProLeuProProArgAsnValG1	793
Db	1260	GCCCAGGCTGCGAGCCCGCAGCACCCAGCGCCCCCGCTTCTCCACGGAATGTGG	1319
Qy	793	yLysValGlnThrAlaSerSerAlaAsnThrLeuTyrLysThrAsnSerValSerValAs	813
Db	1320	CAAAAGTTTCAGACAGCTCTCTGCTGTAACACCTGTGGAGAACAAACTCTGTAAAGTGGGA	1379
Qy	813	pGlyGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProPr	833
Db	1380	CGGTGGAAGCGCGCAGCGATCTTCTGTCAGATCCGCCAGCTGTCCATCCACCGCTGCCCC	1439
Qy	833	oLeuArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysTh	853
Db	1440	TCCTTCGCGTGACATCTACCAATCCCTGACCCCGCAGCGCCCCCGCTTGCACAGAC	1499
Qy	853	rProSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerG1	873
Db	1500	GCCCGAGGTAAAGAGCTTTGAGCCAGCCAGCAGCAGCTGCCCGCTGGGATCTCACA	1559
Qy	873	nileArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAl	893
Db	1560	GATCAGGCCCCCCTCTGCCCCCAGCAGCGCCCGCTCCCGCAGAGAAGAGCCTGC	1619
Qy	893	aPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValA	913
Db	1620	GCCAGGGGCT-GACAAGTCCACCCCTGACCAACAAAGGCGCAACCCGAGAGGACTGTGG	1678

QY 913 spLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProA 933
 Db 1679 ATCTCTGCAACGGAAGCTCTGGGTCTCTGTCTCAATGCTATGCTCTGCAGCCCTG 1738
 QY 933 laProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuT 953
 Db 1739 CACCATCTCTAGGAAGTCGAGGACCAAGTTGAGCCCTAAGCGGTGAAGCGCTCT 1798
 QY 953 yrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValLysLeuV 973
 Db 1799 ATAACGTGTGGTGCACACCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCG 1858
 QY 973 alAspGlyGluAlaAspGlnGluThrTrpLleGlyHisIleAspGlyAspProGlyArgL 993
 Db 1859 TCGACGGGGAGGAGGACGAGGAGTGTGGATTGGCCACATTGATGAGATCTCTGTGCGCA 1918
 QY 993 ySgLYAlaPheProValSerPheValHisPheIleAlaAsp 1006
 Db 1919 AAGGCCATCTCCGGTGTCAATTGTGCACTTTATCGCTGAC 1959

RESULT 2

ID ADB62160 standard; cDNA; 3123 BP.

AC ADB62160;

XX 04-DEC-2003 (first entry)

XX Human cDNA encoding clone CTONG20004520.

XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy; tissue regeneration; cell regeneration; membrane protein; signal transduction-related protein; transcription-related protein; osteoporosis; neurological disease; cancer; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 146..772

FT /*tag= a

FT /product= "Clone CTONG20004520 protein"

XX EP1308459-A2.

XX 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.

XX 05-NOV-2001; 2001JP-00379298.

XX 25-JAN-2002; 2002US-00350978.

XX (HELI-) HELIX RES INST.

PA (RENS-) RES ASSOC BIOTECHNOLOGY.

XX

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S; Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I; Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

DR WPI; 2003-450961/43.

DR P-PSDB; ADB64130.

PT New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.

XX Claim 1; Page; 222pp; English.

XX

CC The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide

CC with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related genes may be included in them, for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy. The genes are involved in tissue and/or cell regeneration. Membrane proteins, signal transduction-related proteins, transcription-related proteins, disease-related proteins and genes encoding them can be used as indicators for diseases (e.g. osteoporosis, neurological diseases, cancer, tumours). The cDNA may be used to regulate the activity or expression of the encoded protein to treat diseases. The sequence presented is a cDNA of the invention. Note: Some of the sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

XX Sequence 3123 BP; 870 A; 671 C; 610 G; 972 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4.63e-98 Length: 3123
 Score: 110.00 Matches: 165
 Percent Similarity: 98.80% Conservatives: 0
 Best Local Similarity: 98.80% Mismatches: 1
 Query Match: 10.93% Indels: 2
 DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x ADB62160 (1-3123)

QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
 Db 272 CCCCTGACCCCGACCGCCGCCCCACCCGTTGCCAAGACGCCAGCGTAAATGGAAGCCTTG 331
 QY 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProLeuPro 880
 Db 332 AGCAGCGGAGCAAGCCTGCCCCCTGGGATCTCACAGATCAGCGCCCCACCTCTGCC 391
 QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerTh 900
 Db 392 CCACAGCGCCCGACCGCGCTCCCGCAGAGAGAGCTGGCGCGGGGCT-GACNAGTCCAC 450
 QY 900 rProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920
 Db 451 CCCACTGACCAACAAGGCGCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCT 510
 QY 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGl 940
 Db 511 GGGTCCTCTGTCCAAATGCTATGGTCCTGACAGCCCCCTGCACCCATGCTTAGGAAGTCGCA 570
 QY 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
 Db 571 GGCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATTAACCTGTGTGGCTGACACCC 630
 QY 960 oAspGluLeuThrPheSerGluGlyAspValLysLeuValAspGlyGluGluAspGlnGl 980
 Db 631 CGATGAGCTCACCTTCTCCGAGCGGGATGTGATCATCTGTGGACGCGGAGGAGGACGAGCA 690
 QY 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
 Db 691 GTGTGGATTGGCCACATTGATGAGATCTCTGGTGCACAAAGCGCATTTCCCGGTGTCAAT 750
 QY 1000 eValHisPheIleAlaAsp 1006
 Db 751 TGTGCACTTTATCGCTGAC 769

RESULT 3

ABK43447

ID ABK43447 standard; cDNA; 3436 BP.

XX ABK43447;

DT 05-JUN-2002 (first entry)

XX DE DNA encoding novel central nervous system protein #27.
XX KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasia; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX OS Homo sapiens.
XX KW WO200155318-A2.
XX PD 02-AUG-2001.
XX XX 17-JAN-2001; 2001WO-US001332.
XX XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226682P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249277P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.

181	CTGTGACCCAAAGCCTTATCTGGAAGATTTAAATTTCTCACGTTACAGTGTGAATATGAATGG	240
Qy		
692	ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro	711
Qy		
241	CGACTACTCCACGAAGACCTGGATGNAAGTGATGACGACATGGATGAGAAATTTGCAGCCC	300
Db		
712	SerGlu-AenArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAenGlnLe	731
Qy		
301	AGTCC-CAACGGCGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTTCCAACCAAGCT	359
Db		
731	uGlnSerAenAlaValSerLeuAlaIaArgApAlaAlaAenLeuAlaLysGlnIa	751
Qy		
360	TCAAGCTAAAGCTGTATCTTTGGCCAGAGATGCTGTCAAAACCTTGGCCAAAGAGAAGCAGAG	419
Db		
751	galaPheMetProSerIleLeuGlnAenGluThrTyrGlyAlaLeuLeuSerGlySerPr	771
Qy		
420	GGCTTTTCATGCCAGCATCTTGCAGATGAGACTTACGGAGCCCTCTCTGAGTGGCAGCCC	479
Db		
771	oProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAs	791
Qy		
480	ACCTCCCGCCAGCCTGACGGCCCCCAGACCACCAGCGCCCCCCCCCGCTTCTCTCCACGGAA	539
Db		
791	nValGlyLys 794	
Qy		

DB	540	IGTTGGCAAA	549
RESULT 4			
ADIS53834			
ID	ADIS53834	standard; cDNA; 3436 BP.	
XX	AC		
XX	AC	ADIS53834;	
XX	XX		
DT	22-APR-2004	(first entry)	
XX	XX		
XX	XX		
DE	DE	cDNA encoding novel human protein seq id 37.	
XX	XX		
KW	KW	neuroprotective; nontropic; antiparkinsonian; anticonvulsant;	
KW	KW	antiadrenergic; antirheumatic; antiarthritic; dermatological;	
KW	KW	antiinflammatory; immunosuppressive; antithyroid; antianemic; vasotropic;	
KW	KW	anti-HIV; hepatotropic; virucide; antibacterial; fungicide;	
KW	KW	antiparasitic; muscular; gynaecological; gastrointestinal; respiratory;	
KW	KW	cardiovascular; antiarteriosclerotic; antiarrhythmic; cardiac;	
KW	KW	nephrotropic; litholytic; cytostatic; gene therapy; neural disorder;	
KW	KW	Alzheimer's disease; Parkinson's disease; Huntington's chorea;	
KW	KW	anyotrophic lateral sclerosis; multiple sclerosis;	
KW	KW	immune system disorder; diabetes; rheumatoid arthritis;	
KW	KW	systemic lupus erythematosus; autoimmune thyroiditis; haemolytic anaemia;	
KW	KW	inflammatory disorder; ischaemia-reperfusion injury;	

KW	inflammatory bowel disease; Crohn's disease; infectious disease;
KW	HIV infection; hepatitis infection; bacterial infection;
KW	fungal infection; parasitic infection; muscular disorder;
KW	reproductive disorder; gastrointestinal disorder; pulmonary disorder;
KW	cardiovascular disorder; atherosclerosis; arrhythmia; myocarditis;
KW	renal disorder; acute glomerulonephritis; pyelonephritis;
KW	renal lithiasis; proliferative disorder; cancerous diseases; human; gene;
ss.	
XX	
XX	
OS	Homo sapiens.
XX	
PN	US2004018969-A1.
XX	
PD	29-JAN-2004.
XX	
XX	
PF	17-JAN-2001; 2001US-00764875.
XX	
PR	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215113P.
 PR 07-JUL-2000; 2000US-02166647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225211P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226688P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 21-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.

PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2004-122079/12.
 DR P-PSDB; ADI54432.
 XX
 PT New polypeptides and nucleic acid molecules, useful for detecting,
 PT preventing, diagnosing, prognosticating, treating or ameliorating medical
 PT conditions e.g. neural disorders, reproductive disorders or infectious
 PT diseases.
 XX
 PS Claim 1; SEQ ID NO 37; 413pp; English.
 XX


```
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250319P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-581633/65.
XX P-PSDB; RAU87438.
XX
XX New isolated nucleic acid encoding a protein for diagnosing, preventing,
XX treating or ameliorating medical conditions and used as food additives or
XX preservatives.
XX
XX Claim 1; SEQ ID NO 358; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX novel central nervous system protein. (I) and polypeptides (II) encoded
XX by (I), are used to treat a medical conditions and in diagnosis of a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angioneurotic nervous system disorders e.g. Alzheimer's disease and
XX amyotrophic lateral sclerosis, infections caused by bacteria, viruses
XX e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
XX e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
XX adenocarcinomas and irritable bowel syndrome, reproductive system
XX disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
XX and pituitary dwarfism, cancers and disorders at the cellular level e.g.
XX leukaemia, disorders involving neovascularisation e.g. malignancies,
XX respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
XX acute kidney failure and blood related disorders e.g. myocardial
XX infarction. The polypeptides can also be used to aid wound healing and
XX epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX maintain organs before transplantation, for supporting cell culture of
XX primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:
Pred. No.: 2,51e-67 Length: 3413
Score: 79.00 Matches: 157
Percent Similarity: 97.52% Conservative: 0
Best Local Similarity: 97.52% Mismatches: 2
Query Match: 7.85% Indels: 4
DB: Gaps: 0

US-09-914-042-1 (1-1006) x ABK43768 (1-3413)

Qy 848 ProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAla 867
Db 572 CCACCCGTTGCCAAGACGCCCGCAGCGTAAATGAAGCCTTGACCGCCAGCGCTGCC 631
Qy 868 ProProGlyIleSerGlnIleArgProProProLeuProProGlnProProSerArgLeu 887
Db 632 CGGCTGGGATCTCACAGATCAGGCCCCACCTCTGCCCCACAGCCGCCCGCCCTC 691
Qy 888 ProGlnLysLysProAlaPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyCl 907
Db 692 CCGCAGAAGAAGCCTGCGCGGGGGCT-GACAAGTCCACCCCTGACCAACAAGGGCA 750
Qy 907 nProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMe 927
Db 751 ACCGAGGAGGACCTGTGGATCTCTTCNACGAGAGCTCTGGGTCTCTGTCATGCTRT 810
Qy 927 t-ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProL 947
```


Db 811 -GGTCCTGCAGCCCTCGACCCCATGCTAGGAAGTCGAGCAACCAAGTTGAAGCCTA 869
Qy 947 ysArgVallysalaleuTyraenCysValAlaAspAsnProAspGluLeuThrPheSerG 967
Db 870 AGCGGCTGAAGCGCTCTATAACTGTGTGGCTGACAAACCCGATGACCTTCCTCCG 929
Qy 967 luGlyAspVallelleValaspGlyGluGluAspGlnGluTrpTrpIleGlyHisIleA 987
Db 930 AGGGGATGTGATCATCTGTCAGCGGGAGGAGGACGAGGTGGTGGATTCGCCACATTTG 989
Qy 987 spGlyAspProGlyArglyGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 990 ATGGAGATCTGTGTCGAAGCGCATTCCTCCGCTGTCATTTGTGCACCTTTATCGCTGAC 1048
RESULT 6
ID ADI54155 standard; cDNA; 3413 BP.
XX ADI54155;
AC ADI54155;
XX
DT 22-APR-2004 (first entry)
XX cDNA encoding novel human protein seq id 358.
XX
KW neuroprotective; neurotropic; antiparkinsonian; anticonvulsant;
KW antidiabetic; antirheumatic; antiarthritic; dermatological;
KW antiinflammatory; immunosuppressive; antithyroid; antianemic; vasotropic;
KW anti-HIV; hepatotropic; virucide; antibacterial; fungicide;
KW antiparasitic; muscular; gynaecological; gastrointestinal; respiratory;
KW cardiovascular; antiarteriosclerotic; antiarrhythmic; cardiant;
KW nephrotropic; litholytic; cytostatic; gene therapy; neural disorder;
KW Alzheimer's disease; Parkinson's disease; Huntington's chorea;
KW amyotrophic lateral sclerosis; multiple sclerosis;
KW immune system disorder; diabetes; rheumatoid arthritis;
KW systemic lupus erythematosus; autoimmune thyroiditis; haemolytic anaemia;
KW inflammatory disorder; ischaemia-reperfusion injury;
KW inflammatory bowel disease; Crohn's disease; infectious disease;
KW HIV infection; hepatitis infection; bacterial infection;
KW fungal infection; parasitic infection; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; pulmonary disorder;
KW cardiovascular disorder; atherosclerosis; arrhythmia; myocarditis;
KW renal disorder; acute glomerulonephritis; pyelonephritis;
KW renal lithiasis; proliferative disorder; cancerous diseases; human; gene;
ss.
XX
OS Homo sapiens.
XX
PN US2004018969-A1.
XX
PD
XX
XX 29-JAN-2004.
XX
PF 17-JAN-2001; 2001US-00764875.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.

PT New polynucleotide sequences obtained from various cDNA libraries, useful
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene
 PT mapping, in the recombinant production of protein, or in generating
 PT antisense DNA or RNA.

XX Claim 1; SEQ ID NO 27010; 44pp; English.

PS The invention relates to an isolated polynucleotide comprising any one of
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
 CC determined by the technique of SBH (sequencing by hybridisation). Also
 CC included is a purified polypeptide comprising a sequence corresponding to
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences
 CC are useful in diagnostics as expressed sequence tags (EST) for
 CC identifying expressed genes or for physical mapping of the human genome,
 CC in forensics, in assessing biodiversity, or in identifying mutations
 CC responsible for genetic disorders and other traits. The nucleotide
 CC sequences are also useful as hybridisation probes, as oligomers for PCR,
 CC for chromosome and gene mapping, in the recombinant production of
 CC protein, or in generating antisense DNA or RNA. The purified polypeptide
 CC is useful for generating antibodies specific for it. The present sequence
 CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030073623

XX Sequence 442 BP; 137 A; 97 C; 101 G; 106 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 3.5e-67 Length: 442
 Score: 78.00 Matches: 78
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.75% Indels: 0
 DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x ACH39798 (1-442)

Qy 365 CysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGlu 384
 Db 38 TGCCTTTGACCTCATTCATGACAGAACTTACCACCTTTCAAGCTGAAGTGAACAGGAA 97
 Qy 385 CysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAlaPhe 404
 Db 98 TGTCAAATATGATGATCTGTGCTGCAAAATAGCAAGAAGAGCTTTAAACAATGCATTT 157
 Qy 405 LysGlyAspAspAsnThrGlyGluAsnAlaValGlnGluLeuThrLysGluIleIle 424
 Db 158 AAGGGGGATGACATCTCTGGAGAAATACATCTGTCAGAACTGACAAAGAGATCATC 217
 Qy 425 SerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaPro 442
 Db 218 TCAGAAATGCAGAGGATGACGGCAATGACGTCTGCTGCTGTGGGGCGCCA 271

RESULT 8

AAV59105
 ID AAV59105 standard; cDNA; 5954 BP.

XX AAV59105;

XX 17-OCT-2003 (revised)

DT 20-NOV-1998 (first entry)

XX Zebrafish differentiation enhancing factor 2 gene.

XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;
 KW C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;
 KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;
 KW nervous system.

OS Danio rerio.

XX Key Location/Qualifiers
 FH 433..3381
 FT CDS

FT /*tag= a
 FT /product= "DEF2 protein"

XX WO9836065-A1.

XX 20-AUG-1998.

XX 13-FEB-1998; 98WO-US002724.

XX 14-FEB-1997; 97US-0038191P.

XX (DAND) DANA FARBER CANCER INST INC.

XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;

XX WPI; 1998-467173/40.

XX P-PSDB; AAW77288.

XX New nucleic acid encoding differentiation enhancing factor - used
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating
 PT tumours and neurological disease.

PS Claim 7; Fig 14; 203pp; English.

XX The differentiation enhancing factors (DEF), comprise at least one each
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
 CC are mediators of SH3-domain dependent signalling and may be involved in
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,
 CC endocytosis or adhesion, migration, proliferation and differentiation of
 CC cells. Typical applications of DEF and agents that modulate interaction
 CC between the protein and its ligand, or of nucleic acid expressing them,
 CC are treatment of hyperplastic and neoplastic disease (a wide range of
 CC solid tumours and leukaemias), including metastases; for in vitro
 CC induction of differentiation of neural crest cells to neurons, glial
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in
 CC the nervous system (e.g. treatment of traumatic injury, stroke,
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral
 CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to
 CC standardise OS field)

SQ Sequence 5954 BP; 1731 A; 1382 C; 1372 G; 1469 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.32e-56 Length: 5954
 Score: 68.00 Matches: 68
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.76% Indels: 0
 DB: 2 Gaps: 0

US-09-914-042-1 (1-1006) x AAV59105 (1-5954)

Qy 171 HisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLysGlu 190
 Db 943 CACGGAATGATCCGACCGAGATCAGCGAGCAGAGATAGCAGAAGAGATGAAAAAGAG 1002
 Qy 191 ArgArgPheGlnLeuGlnMetCysGluTyrLeuLeuValAsnGluIleLysIle 210
 Db 1003 CGGCGTTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAGTCAATGAATCAAGATC 1062
 Qy 211 LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhe 230
 Db 1063 AAAAAAGGTGTCGACCTGCTCCAGAACTCATCAATACTTCCACGCACAGTGCACACTTC 1122
 Qy 231 PheGlnAspGlyLeuLysAlaVal 238
 Db 1123 TTTCAGGATGCTCTCAAAAGCGGTG 1146

RESULT 9

AAV58478

ID AAV58478 standard; cDNA; 563 BP.

XX

```

AC AAS58478;
XX
XX
XX 13-FEB-2002 (first entry)
XX
DE cDNA #1154 encoding portion of a human colon tumour protein.
XX
XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.
XX
XX Homo sapiens.
XX
XX WO200173027-A2.
XX
XX PD 04-OCT-2001.
XX
XX PF 22-MAR-2001; 2001WO-US009246.
XX
XX PR 24-MAR-2000; 2000US-0191597P.
XX
XX PR 04-MAY-2000; 2000US-0202024P.
XX
XX PR 05-MAY-2000; 2000US-0202189P.
XX
XX (CORI-) CORIXA CORP.
XX
XX PI Meagher MJ, Xu J, King GE;
XX
XX DR WPI; 2001-611627/70.
XX
XX PT New colon tumor proteins and related nucleic acid, useful for treatment,
XX PT prevention, diagnosis and monitoring of cancer.
XX
XX PS Claim 4; Page 238; 299pp; English.
XX
XX Th present invention relates to the isolation of novel cDNA sequences
XX encoding for at least an immunogenic portion of human colon tumour
XX proteins. The sequences of the invention are useful in pharmaceutical
XX compositions and vaccines for the prevention and treatment of cancers
XX such as colon cancer. They are also useful for the diagnosis and
XX monitoring of such cancers. Antibodies to the colon tumour proteins and
XX antigen presenting cells that express polynucleotides encoding colon
XX tumour proteins can be used to inhibit the development of cancers. T-
XX cells that react specifically with colon tumour proteins are useful for
XX removing tumour cells from samples (e.g. blood) and for cancer treatment.
XX The polynucleotide sequences are also useful in gene therapy. AAS57325-
XX AAS5880 represent the cDNA sequences of the invention that encode for
XX portions of human colon tumour proteins
XX
XX SQ Sequence 563 BP; 163 A; 131 C; 140 G; 119 T; 0 U; 10 Other;

Alignment Scores:
Pred. No.: 2,248-37 Length: 563
Score: 48.00 Matches: 131
Percent Similarity: 96.32% Conservative: 0
Best Local Similarity: 96.32% Mismatches: 2
Query Match: 4.77% Indels: 5
DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x AAS58478 (1-563)

QY 401 AenAenAlaPheLysGlyAspAspAenThrGlyGluAenAenIleValGlnGluLeuThr 420
DB 2 AACATGCAATTTAAGGGGGATGACAATACTGGAGAAATAATCATCGTCCAGAACTGACA 61
QY 421 LysGluIleIleSerGluValGln-ArgMetThrGlyAenAenValCysCysAspCysG1 440
DB 62 AAGGAGATCATCTCAGAAGTGCT-GAGGATGACGGGCAATGACGTCTGCTGTGACTGTGG 120
QY 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
DB 121 GCGCGCAGATCTCATATGCTTTCCACCAACCTGGGCATCTCGACCTGCATCGATGTTTC 180
QY 460 rGlyIleHisArgGluLeuGlyValHisTrpSerPro-MetGlnSerLeuThrLeuAspV 480
DB 181 CGGAATCCACCAGAGCTGGGGGTTTCATTATTATCCAG-GATGCGATCCCTGACCTTAGATG 239

```

```

QY 480 allLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500
DB 240 TACTGGGACATCTGAGCTGCTGCTGCCAAGATATTTGGGAATGCAGGCTTTAATGAGA 299
QY 500 leMetGluCysCysLeuProAlaGluAsePserValLysProAsnProGlySer-AspMet 519
DB 300 TCATGGAATGTGCTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGGAGCCGACATG 359
QY 520 AenAlaArgLysAspTyrIleThrAlaLysGlyIleGluArg 533
DB 360 AATGCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGG 401

RESULT 10
ID AAV59106 standard; cDNA; 4595 BP.
XX
AC AAV59106;
XX
XX 17-OCT-2003 (revised)
XX 20-NOV-1998 (first entry)
XX
XX Zebrafish differentiation enhancing factor 3 gene.
XX
XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;
XX C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;
XX adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;
XX nervous system.
XX
XX Danio rerio.
XX
XX Key Location/Qualifiers
XX CDS 300..3011
XX /*tag= a
XX /product= "DEF3 protein"
XX
XX WO9836065-A1.
XX
XX 20-AUG-1998.
XX
XX 13-FEB-1998; 98WO-US002724.
XX
XX 14-FEB-1997; 97US-0038191P.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;
XX
XX WPI; 1998-467173/40.
XX P-PSDB; AAW59106.
XX
XX New nucleic acid encoding differentiation enhancing factor - used
XX particularly to regulate adipogenesis and neurogenesis, e.g. for treating
XX tumours and neurological disease.
XX
XX Claim 7; Fig 15; 203pp; English.
XX
XX The differentiation enhancing factors (DEF), comprise at least one each
XX of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
XX domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
XX are mediators of SH3-domain dependent signalling and may be involved in
XX cellular gene expression, cytoskeletal architecture, protein trafficking,
XX endocytosis or adhesion, migration, proliferation and differentiation of
XX cells. Typical applications of DEF and agents that modulate interaction
XX between the protein and it's ligand, or of nucleic acid expressing them,
XX are treatment of hyperplastic and neoplastic disease (a wide range of
XX solid tumours and leukemias), including metastases; for in vitro
XX induction of differentiation of neural crest cells to neurons, glial
XX cells etc.; for increasing neuron survival, and inducing cell repair, in
XX the nervous system (e.g. treatment of traumatic injury, stroke,
XX Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral
XX sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to
XX standardise OS field)

```

SQ Sequence 4595 BP; 1378 A; 1014 C; 1030 G; 1173 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.15e-20 Length: 4595
 Score: 32.00 Matches: 32
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.18% Indels: 0
 DB: 2 Gaps: 0

US-09-914-042-1 (1-1006) x AAV59106 (1-4595)

QY 125 PheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLys 144

DB 672 TTTCACATAGACAGCTCTCTAAAGGAGACCTCAAGGAGTCAAGGGGATTTGAAAAG 731

QY 145 ProPheAspLysAlaTrpLysAspTyrGluThrLys 156

DB 732 CCATTTGATTAAGCATGGAAGGATTTGAAACAAA 767

RESULT 11

ABA08628/c

ID ABA08628 standard; cDNA; 409 BP.

AC ABA08628;

XX 11-JAN-2002 (first entry)

XX Human GTPase activating protein homologue-encoding cDNA, SEQ ID NO:404.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cyostatic; osteopathic; vasotropic; cardiac; virucide; antibacterial;
 KW antifungal; vulnery; antiulcer; ss.

XX Homo sapiens.

OS

XX WO200157188-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US003800.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

XX P-PSDB; ABB11384.

XX Human proteins and DNA encoding sequences useful for preventing, treating
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
 PT and cancer.

XX Claim 1; Page 485; 1963pp; English.

XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA08574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides

CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis; cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness.
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a cDNA encoding a
 CC novel human polypeptide of the invention

SQ Sequence 409 BP; 98 A; 107 C; 74 G; 130 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.1e-19 Length: 409
 Score: 30.00 Matches: 30
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.98% Indels: 0
 DB: 4 Gaps: 0

US-09-914-042-1 (1-1006) x ABA08628 (1-409)

QY 127 LeuAspSerLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPhe 146

DB 133 TTGGATTCCTTTGTTAAAGGAGACCTAAAGGAGTCAAGGAGATCTCAAGAGGCCATTT 74

QY 147 AspLysAlaTrpLysAspTyrGluThrLys 156

DB 73 GACAAAGCCTGGAAGACTATGAGACAAG 44

RESULT 12

ADD71214

ID ADD71214 standard; cDNA; 3775 BP.

XX AC ADD71214;

XX 15-JAN-2004 (first entry)

XX Human intracellular signalling molecule INTS1G-51 cDNA SEQ ID NO:103.

XX human; intracellular signalling molecule; INTS1G; cyostatic;
 KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;
 KW neuroprotective; cerebroprotective; anti-HIV; antiatherogenic;
 KW antiinflammatory; thymimetic; gene therapy;
 KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;
 KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;
 KW immune disorder; inflammatory disorder; AIDS; allergy;
 KW developmental disorder; Hypothyroidism; Cushing's syndrome;
 KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.

```

OS Homo sapiens.
XX WO2003039348-A2.
XX PD 15-MAY-2003.
XX PF 07-NOV-2002; 2002WO-US036151.
XX PR 09-NOV-2001; 2001US-0344472P.
XX PR 30-NOV-2001; 2001US-0344472P.
XX PR 14-DEC-2001; 2001US-0340296P.
XX PR 21-DEC-2001; 2001US-0343557P.
XX PR 18-JAN-2002; 2002US-0350420P.
XX PR 25-JAN-2002; 2002US-0351927P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Kable AE, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;
PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;
PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;
PI Marquis JP, Gietzen KJ, Baughn MR, Borowsky ML, Yao MG, Chawla NK;
PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;
PI Lee SY, Tran UK, Elliott VS, Sprague WM, Tang YT, Zebajadian Y;
PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;
PI Zheng W;
XX DR WPI; 2003-441441/41.
XX DR P-PSDB; ADD71162.
XX PT New human intracellular signalling molecules (INTSIG)), useful for
XX PT diagnosing, treating and preventing diseases or conditions associated
XX PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or
XX PT infections.
XX PS Claim 5; SEQ ID NO 103; 363pp; English.
XX CC The present invention describes human intracellular signalling molecules
XX CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins
XX CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,
XX CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
XX CC antiinflammatory and thymimetic activities, and can be used in the
XX CC therapy. The INTSIG polynucleotides and proteins can be used in the
XX CC diagnosis, treatment and prevention of diseases or conditions associated
XX CC with the decreased expression or overexpression of INTSIG, such as cell
XX CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),
XX CC neurological (e.g. epilepsy, Huntington's disease, stroke),
XX CC immune/inflammatory (e.g. AIDS, allergies), reproductive and vesicle-trafficking
XX CC disorders, or infections. They are also useful in assessing the effects
XX CC of exogenous compounds on the expression of nucleic acid and amino acid
XX CC sequences of INTSIG. The INTSIG or its fragments are useful in screening
XX CC compounds for effectiveness as agonist or antagonist of the polypeptides,
XX CC or in altering the expression of the target polynucleotide and compounds
XX CC that specifically bind to or modulate the activity of the polypeptide.
XX CC The present sequence encodes human INTSIG-51 from the present invention.
XX SQ Sequence 3775 BP; 1123 A; 989 C; 873 G; 790 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9,136-19 Length: 3775
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.98% Indels: 0
DB: 10 Gaps: 0

US-09-914-042-1 (1-1006) x ADD71214 (1-3775)
Qy 127 LeuApsSerLeuLeuLeuGlyAspLeuLeuGlyValLeuGlyAspLeuLeuGlyProphe 146
Db 696 TTGGATTCTTTGTTAAAGGAGACCTAAAGGGAGTCAAGGAGATCTCAAGAGGCATTT 755
Qy 147 AspLeuAlaTrpLysAspTyrGluThrLys 156

```

Db 756 GACAAGCCTGGAAGATTATGACACAAAG 785

RESULT 13

ADD71207

ID ADD71207 standard; cDNA; 3946 BP.

XX AC ADD71207;

XX DT 15-JAN-2004 (first entry)

XX DE Human intracellular signalling molecule INTSIG-44 cDNA SEQ ID NO:96.

XX KW human; intracellular signalling molecule; INTSIG; cytostatic;

XX KW antiarteriosclerotic; antidiabetic; anticonvulsant; neurotropic;

XX KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;

XX KW antiinflammatory; thymimetic; gene therapy;

XX KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;

XX KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;

XX KW immune disorder; inflammatory disorder; AIDS; allergy;

XX KW developmental disorder; Hypothyroidism; Cushing's syndrome;

XX KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.

XX OS Homo sapiens.

XX PN WO2003039348-A2.

XX PD 15-MAY-2003.

XX PF 07-NOV-2002; 2002WO-US036151.

XX PR 09-NOV-2001; 2001US-0344472P.

XX PR 30-NOV-2001; 2001US-0344472P.

XX PR 14-DEC-2001; 2001US-0340296P.

XX PR 21-DEC-2001; 2001US-0343557P.

XX PR 18-JAN-2002; 2002US-0350420P.

XX PR 25-JAN-2002; 2002US-0351927P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Kable AE, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;

XX PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;

XX PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;

XX PI Marquis JP, Gietzen KJ, Baughn MR, Borowsky ML, Yao MG, Chawla NK;

XX PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;

XX PI Lee SY, Tran UK, Elliott VS, Sprague WM, Tang YT, Zebajadian Y;

XX PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;

XX PI Zheng W;

XX DR WPI; 2003-441441/41.

XX DR P-PSDB; ADD71155.

XX PT New human intracellular signalling molecules (INTSIG)), useful for

XX PT diagnosing, treating and preventing diseases or conditions associated

XX PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or

XX PT infections.

XX PS Claim 5; SEQ ID NO 96; 363pp; English.

XX CC The present invention describes human intracellular signalling molecules

XX CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins

XX CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,

XX CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,

XX CC antiinflammatory and thymimetic activities, and can be used in the

XX CC therapy. The INTSIG polynucleotides and proteins can be used in the

XX CC diagnosis, treatment and prevention of diseases or conditions associated

XX CC with the decreased expression or overexpression of INTSIG, such as cell

XX CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),

XX CC neurological (e.g. epilepsy, Huntington's disease, stroke),

XX CC immune/inflammatory (e.g. AIDS, allergies), reproductive and vesicle-trafficking

XX CC disorders, or infections. They are also useful in assessing the effects

XX CC of exogenous compounds on the expression of nucleic acid and amino acid

CC sequences of INTSIG. The INTSIG or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The present sequence encodes human INTSIG-44 from the present invention.

XX SQ Sequence 3946 BP; 1165 A; 1064 C; 902 G; 815 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	9.52e-19	Length:	3946
Score:	30.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.98%	Indels:	0
DB:	10	Gaps:	0

US-09-914-042-1 (1-1006) x ADD71207 (1-3946)

QY	127	LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPhe	146
DB	696	TGGATTCTTTGTTAAAGGAGACCTAAAGGAGTCAAAGGAGATCTCAAGAGCCATT	755
QY	147	AspLysAlaTyrLysAspTyrGluThrLys	156
DB	756	GACAAAGCTGGAAAGATTATGAGACAAAG	785

RESULT 14

ADJ51008
ID ADJ51008 standard; cDNA; 3974 BP.

XX AC ADJ51008;

XX DT 06-MAY-2004 (first entry)

XX DE Human cDNA encoding NOV31a.

XX KW Human; ss; gene; NOVX; autoimmune disease; Alzheimer's disease; stroke;
KW allergy; Parkinson's disease; Huntington's disease; multiple sclerosis;
KW anxiety; pain; diabetes; graft versus host disease; pancreatitis;
KW obesity; ulcer; anaemia; cancer; viral infection; bacterial infection;
KW parasitic infection.

XX OS Homo sapiens.

XX PN US2004030096-A1.

XX PD 12-FEB-2004.

XX PF 01-AUG-2002; 2002US-00210281.

XX PR 02-AUG-2001; 2001US-0309501P.

XX PR 03-AUG-2001; 2001US-0310291P.

XX PR 08-AUG-2001; 2001US-0310951P.

XX PR 09-AUG-2001; 2001US-0311292P.

XX PR 13-AUG-2001; 2001US-0311979P.

XX PR 14-AUG-2001; 2001US-0312203P.

XX PR 17-AUG-2001; 2001US-0313201P.

XX PR 20-AUG-2001; 2001US-0313643P.

XX PR 20-AUG-2001; 2001US-0313702P.

XX PR 21-AUG-2001; 2001US-0314031P.

XX PR 23-AUG-2001; 2001US-0314466P.

XX PR 28-AUG-2001; 2001US-0315403P.

XX PR 29-AUG-2001; 2001US-0315853P.

XX PR 05-MAR-2002; 2002US-0361775P.

XX PR 05-MAR-2002; 2002US-0361832P.

XX (GORM/) GORMAN L.

PA (ZERH/) ZERHUSEN B D.

PA (EDIN/) EDINGER S R.

PA (PADI/) PADIGARU M.

PA (GUOX/) GUO X.

PA (KEKU/) KEKUDA R.

PA (ZHON/) ZHONG M.

PA (PATT/) PATTURAJAN M.
PA (MILL/) MILLER C E.
PA (JIWW/) JI W.
PA (PENA/) PENA C E A.
PA (BURG/) BURGESS C E.
PA (SCIO/) SCIORE P.
PA (STON/) STONE D J.
PA (TAUP/) TAUPIER R J.
PA (CASM/) CASMAN S J.
PA (ROTH/) ROTHENBERG M E.
PA (MALY/) MALYANKAR U M.
PA (BOLD/) BOLDOG F L.

Gorman L, Zerhusen BD, Edinger SR, Padigaru M, Guo X, Kekuda R,
Zhong M, Patturajan M, Miller CE, Ji W, Pena CE, Burgess CE,
Sciore P, Stone DJ, Taupier RJ, Casman SJ, Rothenberg ME,
Malyankar UM, Boldog FL;

XX WPI; 2004-168942/16.
DR P-PSDB; ADJ51009.

XX New NOVX polypeptides and polynucleotides, useful in diagnosing, treating
PT or preventing diseases or conditions, e.g. autoimmune disease, treating
PT Alzheimer's disease, diabetes, graft versus host disease, cancer or viral
PT or bacterial infections.

PS Claim 20; SEQ ID NO 73; 342pp; English.

XX The invention relates to an isolated NOVX polypeptide (of 44 disclosed)
CC comprising its mature form, a sequence having at least 98% sequence
CC identity to NOVX or a sequence comprising one or more conservative
CC substitutions in the amino acid sequence of NOVX. Also included are a
CC composition comprising NOVX and a carrier, a kit comprising, in one or
CC more containers, the composition, a method of identifying an agent that
CC binds to NOVX, a method for identifying a potential therapeutic agent for
CC use in treatment of a pathology related to aberrant expression or
CC aberrant physiological interactions of NOVX, a method for screening for a
CC modulator of activity of or of latency or predisposition to a pathology
CC associated with NOVX, a method for modulating the activity of NOVX, a
CC method of treating or preventing a pathology associated with NOVX or a
CC pathological state in a mammal, an isolated nucleic acid molecule
CC encoding a NOVX protein, a vector comprising the nucleic acid molecule,
CC a cell comprising the vector, an antibody that immunospecifically binds
CC to NOVX, a method for determining the presence or amount of NOVX or the
CC nucleic acid molecule in a sample, a method for determining the presence
CC of or predisposition to a disease associated with altered levels of
CC expression of NOVX or the nucleic acid molecule in a first mammalian
CC subject and a method of producing NOVX (comprising culturing the cell
CC under conditions that lead to expression of the polypeptide). NOVX is
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease associated with NOVX. The polypeptides
CC and nucleic acid molecules are useful in diagnosing, treating or
CC preventing diseases or conditions, e.g. autoimmune disease, Alzheimer's
CC disease, stroke, allergies, Parkinson's disease, Huntington's disease,
CC multiple sclerosis, anxiety, pain, diabetes, graft versus host disease,
CC pancreatitis, obesity, ulcers, anaemia, cancer, viral or bacterial and
CC parasitic infections (many more diseases and disorders are listed in the
CC specification). The present sequence encodes a NOVX protein.

XX SQ Sequence 3974 BP; 1208 A; 1019 C; 888 G; 859 T; 0 U; 0 Other;

Alignment Scores:

Pred No.:	9.59e-19	Length:	3974
Score:	30.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	2.98%	Indels:	0
DB:	12	Gaps:	0

US-09-914-042-1 (1-1006) x ADJ51008 (1-3974)

QY 127 LeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPhe 146
|||||

699	TTGGATTCTTTTGTAAAGGAGACCTAAAGGGAGTCAAGGAGATCTCAAGAAGCCATT	755
147	ASPLYSALATRLPYEASPTyrGluThrLys	156
759	GACAAAGCCTGGAAAGATTATGAGACAAAG	788
<p>RESULT 15</p> <p>AAV59104 standard; cDNA; 4382 BP.</p> <p>AAV59104;</p> <p>AC</p> <p>XX</p> <p>17-OCT-2003 (revised)</p> <p>20-NOV-1998 (first entry)</p> <p>XX</p> <p>Zebrafish differentiation enhancing factor 1 gene.</p> <p>XX</p> <p>ss, Zebrafish; differentiation enhancing factor; ankyrin repeat;</p> <p>C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;</p> <p>KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;</p> <p>KW nervous system.</p> <p>XX</p> <p>XX</p> <p>OS Danio rerio.</p> <p>XX</p> <p>Key Location/Qualifiers</p> <p>CDS 351..3806</p> <p>FT /*tag= a</p> <p>FT /product= "DEF1 protein"</p> <p>XX</p> <p>WO9836065-A1.</p> <p>XX</p> <p>20-AUG-1998.</p> <p>XX</p> <p>13-FEB-1998; 98WO-US002724.</p> <p>XX</p> <p>14-FEB-1997; 97US-0038191P.</p> <p>XX</p> <p>(DAND) DANA FARBER CANCER INST INC.</p> <p>XX</p> <p>Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;</p> <p>XX</p> <p>WPI; 1998-467173/40.</p> <p>DR</p> <p>P-PSDB; AAW77287.</p> <p>XX</p> <p>New nucleic acid encoding differentiation enhancing factor - used</p> <p>PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating</p> <p>PT tumours and neurological disease.</p> <p>XX</p> <p>Claim 7; Fig 13; 203pp; English.</p> <p>XX</p> <p>The differentiation enhancing factors (DEF), comprise at least one each</p> <p>CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology</p> <p>CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they</p> <p>CC are mediators of SH3-domain dependent signalling and may be involved in</p> <p>CC cellular gene expression, cytoskeletal architecture, protein trafficking,</p> <p>CC endocytosis or adhesion, migration, proliferation and differentiation of</p> <p>CC cells. Typical applications of DEF and agents that modulate interaction of</p> <p>CC between the protein and it's ligand, or of nucleic acid expressing them,</p> <p>CC are treatment of hyperplastic and neoplastic disease (a wide range of</p> <p>CC solid tumours and leukaemias), including metastases; for in vitro</p> <p>CC induction of differentiation of neural crest cells to neurons, glial</p> <p>CC cells etc.; for increasing neuron survival, and inducing cell repair, in</p> <p>CC the nervous system (e.g. treatment of traumatic injury, stroke, glial</p> <p>CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral</p> <p>CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to</p> <p>CC standardise OS field)</p> <p>XX</p> <p>Sequence 4382 BP; 1222 A; 1132 C; 1098 G; 930 T; 0 U; 0 Other;</p>		
<p>Alignment Scores:</p> <p>Pred. No.: 1.05e-18 Length: 4382</p> <p>Score: 30.00 Matches: 30</p> <p>Percent Similarity: 100.00% Conservative: 0</p>		

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 14:05:47 ; Search time 10576 Seconds

(without alignments)
4609.110 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAFVSPVHFIAID 1006

Scoring table:

OLIGO Xgapop 60.0 Xgapext 60.0
Ygapop 60.0 Ygapext 60.0
Fgapop 6.0 Fgapext 7.0
Delop 6.0 Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9408497

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09914042/Runat_04082005_090746_2657/app_query.fasta_1.1159
-DB=GenEmbl -QWMT=fastcap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914042 @CGN 1 1 6279 @runat_04082005_090746_2657 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=110 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	717	71.3	5711	9	AB007860 Homo sapi
2	525	52.2	5544	9	BC063308 Homo sapi
3	297	29.5	4486	6	I74317 Sequence 15
4	257	25.5	4328	6	I74314 Sequence 12

5	161	16.0	4713	10	AK172944	AK172944 Mus muscu
6	161	16.0	5072	10	BC080847	BC080847 Mus muscu
7	130	12.9	2109	6	CQ720712	CQ720712 Sequence
8	110	10.9	3123	6	AX746789	AX746789 Sequence
9	110	10.9	3123	9	AK091121	AK091121 Homo sapi
10	110	10.9	3870	9	AK124900	AK124900 Homo sapi
11	108	10.7	858	5	CR385625	CR385625 Gallus ga
12	81	8.1	171653	9	AC080162	AC080162 Homo sapi
13	71	7.1	2157	9	AK123185	AK123185 Homo sapi
14	68	6.8	2949	6	AR243296	AR243296 Sequence
15	68	6.8	5954	6	AR243295	AR243295 Sequence
16	55	5.5	158388	2	AC102852	AC102852 Mus muscu
17	55	5.5	241137	2	AC112078	AC112078 Rattus no
18	55	5.5	269229	2	AC097551	AC097551 Rattus no
19	49	4.9	85411	9	AC079782	AC079782 Homo sapi
20	49	4.9	198935	2	AC102817	AC102817 Mus muscu
21	49	4.9	241137	2	AC112078	AC112078 Rattus no
22	48	4.8	563	6	AX261503	AX261503 Sequence
23	46	4.6	3989	5	BC070750	BC070750 Xenopus l
24	42	4.2	191827	9	AC093904	AC093904 Homo sapi
25	37	3.7	1721	5	CR387820	CR387820 Gallus ga
26	34	3.4	1423	10	BC028949	BC028949 Mus muscu
27	34	3.4	65282	2	AC116865	AC116865 Mus muscu
28	34	3.4	177850	2	AC113090	AC113090 Mus muscu
29	32	3.2	2712	6	AR243298	AR243298 Sequence
30	32	3.2	4595	6	AR243297	AR243297 Sequence
31	30	3.0	706	5	BX950804	BX950804 Gallus ga
32	30	3.0	2583	6	CQ716704	CQ716704 Sequence
33	30	3.0	3456	6	AR243294	AR243294 Sequence
34	30	3.0	4148	10	AK122477	AK122477 Mus muscu
35	30	3.0	4150	10	AF075462	AF075462 Mus muscu
36	30	3.0	4370	10	AF075461	AF075461 Mus muscu
37	30	3.0	4382	6	AR243293	AR243293 Sequence
38	30	3.0	4434	10	BC048818	BC048818 Mus muscu
39	30	3.0	5330	4	AF112886	AF112886 Bos tauru
40	30	3.0	5330	5	CR243292	CR243292 Sequence
41	30	3.0	209698	5	CR381585	CR381585 Zebrafish
42	27	2.7	162648	5	AL935033	AL935033 Zebrafish
43	27	2.7	212604	5	AL805945	AL805945 Zebrafish
44	24	2.4	884	5	CF523567	CF523567 Gallus ga
45	24	2.4	171767	10	AC134601	AC134601 Mus muscu

ALIGNMENTS

RESULT 1	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.	5711 bp	linear	PRI 10-JAN-2004
AB007860	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
LOCUS	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
DEFINITION	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
ACCESSION	AB007860	Homo sapiens KIAA0400 mRNA, partial cds.			
VERSION	AB007860.1	GI:2662080			
KEYWORDS					
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
		Ishikawa, K., Nagase, T., Nakajima, D., Seki, N., Ohira, M.,			
		Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.			
TITLE		Prediction of the coding sequences of unidentified human genes.			
		VIII. 78 new cDNA clones from brain which code for large proteins			
		in vitro			
JOURNAL		DNA Res. 4 (5), 307-313 (1997)			
MEDLINE		98116555			
PUBMED		9455477			
REFERENCE		2 (bases 1 to 5711)			
AUTHORS		Chara, O.			
TITLE		Direct Submission			
JOURNAL		Submitted (06-OCT-1997) Osamu Ohara, Kazusa DNA Research Institute,			
		Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba			
		292-0812, Japan (E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913,			
		Fax:+81-438-52-3914)			
FEATURES		Location/Qualifiers			

source

```

1. 5711
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HG1091"
/sex="male"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1. 5711
/gene="KIAA0400"
<293...3361
/gene="KIAA0400"
/note="Start codon is not identified"
/codon_start=1
/protein_id="BAA23696.2"
/db_xref="GI:40788243"
/translation="GGGAPAAVRQPRRAEAMPDQTSVSEFVAETHEDYKAPTASSFTT
RTACRNTVAIAIEALDVRMVLKYMKSVAIKNSGLAHVENEBOYTOALEKFGGNC
VCRDPDLSGSAFLKFSVFTKELTALFKNLIOMNNIISFPLDSLKGLDKGVKGLKK
PFDAKWDYETIKPIEKEKEHAKLHGMIKTEISGAIEAEMEKERRFFLOQCEYL
LKVNEIKIKGVLDLQNLKIYPHACQNFQDGLKAVESLKSIEITLSDTLHIKOQAD
BERQLIOLRDLKSAQVEQKDSQIRKSTAYSLHQPQGNKEHGTENRGLYKSDG
IRKWKQKCSVKQGLFTISHGTANRPAPKLNLTQVKTNPEKKKCFDLISHDRTH
FOADEQCOQIWMVLQNSKEBALNNAFKGDNDGNNI VOELTKETIIEVORTGND
VCCDGPDPWLSNLGILTCIECSGIHRELGVHYSRMQSLTLDVLGTSLLAKNI
GNAFNEIMECLPAEDSVKPNPGSDMNARKDYITAKYIERRYARKHADNAAKHSL
CEAVNRDIFGLQADQADVLTEKIPLANGHEPDETHALHVAHSVORTSLHI VDFLV
QKNGLDRTQKGTSTALHYCCTDNAECLKULLRGKASIEIANESGETPLDIKRLKH
EHCBELTQALSGFNSHVHVEYEWRLHEDLDESDDMDKLOPSPNRRREDRPISEY
OLGNSQLONSAVSLRANALAKKQRAFMPISILQNETYGALLSGSPPAOPAPSTT
SAPLPNPPVAKTQVASSANTLWKTNSVSDGSGRSRSSDPPAVHPPLPLRVLTSTN
PLTPPBPVAKTSVMEALSQSPKAPPGISQIRPPPLPQPPSRPLQKPKAPGADK
STPLNKGQPRGPDLSLATEALGPLSNANWLQPPAPMPKRSQATKLKPRVKALYNV
ADNPDELTFSEGDVIIVDGEEDQWMIHIDGDFGRKGAFFVSVFHTAD"

```

gene

CDS

Qy	141	AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle	160
Db	761	GATCTGAAAAAGCCCTTTTGATAAAAGCTTGGAAAGACTATGAAACAAAAAATACCAAGATA	820
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly	180
Db	821	GAAGAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAATTAAGCGGA	880
Qy	181	AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu	200
Db	881	CGCGAAATTCGGAAGAGATGGAAGAGAGAGGCGCTCTTCCAGCTACAGATGTGCGAG	940
Qy	201	TyrLeuLeuLysValAsnGluLysIleLysLysGlyValAspLeuLeuGlnAsnLeu	220
Db	941	TATCTGCTGAAGGTCACGAAATCAAGATTAAAAAGGGAGTAGATTTACTTTCAGAAATCTG	1000
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	1001	ATCAATACTTTCATGCCAATGCATTTTTTTCAGGATGGACTCAAAAGCCGTGGNAAGC	1060
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1061	CTCAACCTTCCATTGAAACGCTGCTACGGATCTTACACGATCTTCAACAGCCAGGAT	1120
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuValGlu	280
Db	1121	GAAGAAAGAGGCGAGTGTATACAGCTTCGAGATATTTTGAATTCGCAATTCGAGTTGAA	1180
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1181	CAGAAAGGAGNCTCCCAATTCGTACAGGACACAGCTTATAGCTTACATCAGCCTCAGGA	1240
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1241	AACAAGAACATGGGACCGAGCGGAGCGGAGCTCTTACAAGAAAGAGTGCAGGATCCGA	1300
Qy	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1301	AAAGTGTGGCAGAAAAGGAAATGTTTCAAGTTAAAAATGGTTTTTCTGACCATATCCATGGT	1360
Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1361	ACCGCTAACCGGCTCTCGCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCT	1420
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1421	GAGGAGAAGAGTGTCTTGACCTTATTTTCAATGACAGAACTTACCACCTTTTCAAGCTGAA	1480
Qy	381	AspGluGlnCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluLysAlaLeu	400
Db	1481	GATGAAAGGAGATGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAGAAAGAGCTTTA	1540
Qy	401	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1541	AACAATGCATTTAAGGGGATGACAATCTCGGAGAAATAACATCGTCCAAGAACTGACA	1600
Qy	421	LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1601	AAGGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGGG	1660
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1661	CGCGCAGATCTCATCGGCTTTCCACCAACCTGGGATCTTCGACTGCATCGATGTGTTCC	1720
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspVal	480
Db	1721	GGATCCACCGAGAGCTGGGGGTTTCATTTATTTCCAG-GATGCAGTCCCTGACCTTAGATGT	1779
Qy	480	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1780	ACTGGGAACATCTGAGCTGCTCGCCAAAGAAATTTGGGAAATTCAGGCTTTAATAGAT	1839

ORIGIN

Pred. No.:	0	Length:	5711
Score:	717.00	Matches:	1003
Percent Similarity:	99.41%	Conservative:	0
Best Local Similarity:	99.41%	Mismatches:	3
Query Match:	71.27%	Indels:	6
DB:	9	Gaps:	0
US-09-914-042-1 (1-1006) x AB007860 (1-5711)			
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla	20
Db	341	ATGCCGACACAGATCTCCGTGTGCGAATTCGTGGCCGAGACCCATGAGGACTACAAGCG	400
Qy	21	ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	401	CCCACGGCTCCAGCTTCCACCACCGCAGCGGAGTGCAGGAACTGTGTGGCGCCATC	460
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla	60
Db	461	GAGAGGGCTTTGGACGTGGACCGGATGTTCTTTTACAAAATGAAGAAATCCGTGAAGCA	520
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu	80
Db	521	ATCAACAGCTCTGGGCTGGCTCACGTGGAATAATGAAGACAGTACACCCAGGCTCTGGAG	580
Qy	81	LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys	100
Db	581	AAGTTTGGCGCAACTGTGTATGACAGAGATGACCCAGATTTAGGAAGTGCCTTCTCGAAG	640
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	641	TTCTCAGTGTTCNAAGAGAGTTGACAGCATTTCAAAACCTGATTTCAGATATGAAAC	700
Qy	121	AsnIleLeuSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	701	AACATAATCTCCTTCCCTTTGGACAGTTTGTGTAAGGGGGACCTGAAAGGAGTGAAGGG	760

QY	500	eMetGluCysCysLeuProAlaAspSerVallysProAsnProGlySerAspMetAs	520
Db	1840	CATGGAAATGTTGCTACCAAGCTGAGGACTCAGTCAAAACCAACCCAGGAGGAGATGAA	1899
QY	520	nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysIleHi	540
Db	1900	TGCAAGAAGGACTACATACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGCA	1959
QY	540	sAlaAspAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh	560
Db	1960	CGCGGATAACGGGGCGAAGCTTACAGTCTTTTTCGAGGGCGGTCAAAACGAGAGATATTTT	2019
QY	560	eclyLeuLeuGlnAlaTyrAlaAspClyValAspLeuThrGluLysIleProLeuAlaAs	580
Db	2020	TGGATTGCTCCAAAGCTTATGCTGGTGGATCTTACGGAAATAATCCCACTGGCCAA	2079
QY	580	nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe	600
Db	2080	CGGACATGAGCCGGATGAACCGGCTCCACCTTGCACTGAGTCAGATCGTGGATCGAACCTC	2139
QY	600	rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy	620
Db	2140	TCTTCACATTGTAGACTTTTATGTTTACAGAACAGTGGGAACCTGGATAAACACAGAGGAA	2199
QY	620	sGlySerThrAlaLeuHisTyrCysCysLeuThrAspAlaGluCysLeuLysLeuLe	640
Db	2200	AGGAGACACAGCCCTGCACTACTGCTGCTGACCGACAAATGCGAGTGCCTCAAGTTGCT	2259
QY	640	uLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIl	660
Db	2260	CCTGCGGGGAGGGCTTCATCGATAGCAACAGTCCAGGAGAGACTCCGCTGGACAT	2319
QY	660	eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr	680
Db	2320	TGCCAAGCGCTCAAGCAGCAGCACTGTGAGGAGCTGTGACCCCAAGCTTATCTGGAG	2379
QY	680	gPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGl	700
Db	2380	ATTTAAATCTCAGCTTCAGTTGTAATATGATGGCGACTACTCCACGAAGACCTGGATGA	2439
QY	700	uSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgP	720
Db	2440	ANGATGACGACATGGATGAGAAATGCGCCCAAGTCC-CAACCGCGGGGAAGACCGGC	2498
QY	720	roIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaA	740
Db	2498	CCATCAGCTTCTACAGCTGGCTCCACACAGCTTCACTCAACCGTGTATCTTTGGCA	2558
QY	740	rgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnA	760
Db	2558	GAGATGCTGCAAACTTGCAAGGAGAAGCAGAGGGCTTTATGCCCCAGCATCTTGCAGA	2618
QY	760	enGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProS	780
Db	2618	ATGAGACTTACGAGCCCTCTGATGGAGCGCCACCTCCGCGCCAGGCTTCAGAGCCCA	2678
QY	780	erThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerS	800
Db	2678	GCACACACAGCGCCCCCGCTTCTCCACGGAATGTTGGCAAGTTCAGACAGCTCTCT	2738
QY	800	eAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS	820
Db	2738	CTGCTAACACCTGTGGAAGCAAACTCTGTAAGTGTGAGCGGTGGAAGCGGAGCGAT	2798
QY	820	erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA	840
Db	2798	CTTCCTCAGATCCGCGAGCTGCCATCCACGCTGCCCTCTTCGCGTGACATCTACCA	2858
QY	840	snProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaL	860
Db	2858	ATCCCTGACCCCAAGCGGCCCCCAGCCGTTGCCAAGACGCCCGGTAATGGAAGCT	2918
QY	860	eSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuP	880

Db	2919	TGAGCAGCCGAGCAAGGCTGCCCGCCTGGGATCTCAGATCAGGCCACCTCTGC	2978
QY	880	toProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer	899
Db	2978	CCCCACAGCCGCCAGCCCTCCCGCAGAGAACCTGCGCGGGGCT-GACAAAGTCC	3037
QY	900	ThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAla	919
Db	3038	ACCCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGAAGCT	3097
QY	920	LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer	939
Db	3098	CTGGGTCTCTGTCTCAATGCTATGCTGCGAGCCCTGCAACCTGCTAGGAAGTGC	3157
QY	940	GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValaAlaAspAen	959
Db	3158	CAGGACACCAAGTTGACGCTTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAC	3217
QY	960	ProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGln	979
Db	3218	CCCGATGAGTCACTTCTCCGAGGGGATGTGATCATCTGTCGACGGGAGGAGACCA	3277
QY	980	GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer	999
Db	3278	GAGTGGTGGATTGGCCACATTCATGATGAGATCTGTCGCAAGGGCCATTCCTCGGTGTCA	3337
QY	1000	PheValHisPheIleAlaAsp	1006
Db	3338	TTTGTGCACTTATCGCTGAC	3358
RESULT 2			
LOCUS	BC063308	5544 bp	mRNA linear PRI 23-DEC-2003
DEFINITION	Homo sapiens development and differentiation enhancing factor 2, mRNA (cDNA clone MGC:71731 IMAGE:30345039), complete cds.		
ACCESSION	BC063308		
VERSION	BC063308.1		
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 5544)		
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Hopkins, R.F., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toehiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullan, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smal, J., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 5544)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. Stefan Hansson
 cDNA Library Preparation: Michael Brownstein / Ted Usdin
 Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@axil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 133 Row: h Column: 11
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 4502248.

FEATURES

source

1..5544
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:71731 IMAGE:30345039"
 /tissue_type="Placenta, normal"
 /clone_lib="NIH_MGC_147"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 1..5544
 /gene="DDEF2"
 /note="synonyms: KIAA0400, PAP, SHAG1, PAG3, Pap-alpha"
 /db_xref="LocusID:8853"
 /db_xref="MIM:603817"
 294..3179

gene

CDS

/gene="DDEF2"
 /codon_start=1
 /product="DDEF2 protein"
 /protein_id="AAH63308.1"
 /db_xref="GI:38648787"
 /db_xref="LocusID:8853"
 /db_xref="MIM:603817"
 /translation="MPDOI SVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAATEEAL
 DVDRMVLVYMKSKVKAISNSGLAHVENEQVTALEKEGNGRVCEDDPLDLSALFKES
 VFTKELTALFKNLQNMNNIISFPLDSLKGLDKGVKGLDKKPPDKAKWDYETKITKI
 EKEKKEHAKLHGMIKTEISGAIEAEMEKEKRRFFQLMCEYLLKVKNEIKIKKGVDLQ
 NLIKYFHAQCNPFDGLKAVESLKPSIETLSDLTIKQAQDEERRQLIQRDLKSA
 LQVEKQDSQIRQSTAYSLHQPQGNKEHGTNRNGSLYKKSQGIKRWQKRKCSVKNGF
 LTI SHGTNRPPAKNLTLT'COVNTPEBKCFDLTSHDRTYHFOAEDQECCQIWMSVL
 QNSKEALNNAFKGDDNTGNNI'VOELTKEIISVORMTGNVDVCCDCAFPPTWLSTN
 LGILTCISGTHRELGVHYSRMQSLTLDVIGTSELLAKNIGNAGFNEIWECCLPAAE
 DSVKPNPDSNMWARKDYITAKYIERRYARKHADNAAKLHLSCEAVKTRDIFGLLQAY
 ADGVDLTEKIPLANGHEPDEETALHUVRSVDRSLHIVDFLVQNSGNDLKQTKGKSTA
 LHVCLCTONAEKLLLRGKASIEIANESGETPLDIKRLKHEHCEELLTQALSGRFN
 SHVHVEVRLIHLEDSDDDMDDEKLQPSNRRDRPTISFVQLSGNOLQSNVSLAR
 DAANLAKQKRAFMSI'IQNETY'GALLSGSPPPAAPSPTSPAPLP'PRNVKQDPLT
 FTPPPPVAKTFSVMALQSPKAPFPGISQIRPPELPFPQPSRLQKPKAPGADKSTP
 LTNKQPGRPVDLSATEALGPLSNAMVLOPPAPMPKQSOATKLKPKRYALYNCAVDN
 PELDTFSGSDVIIVDGEEDQEWIIGHIDGDPGRKGAPFVSFVHFIAID"
 1209..1490
 /gene="DDEF2"
 /note="PH; Region: Pleckstrin homology domain"
 /db_xref="CDD:smart00233"
 1554..1898

misc_feature

/gene="DDEF2"
 /note="PH; Region: Pleckstrin homology domain"
 /db_xref="CDD:smart00233"
 1554..1898

misc_feature

/gene="DDEF2"
 /note="ArfGap; Region: Putative GTPase activating protein
 for Arf. Putative zinc fingers with GTPase activating
 proteins (GAPs) towards the small GTPase, Arf. The GAP of
 ARD1 stimulates GTPase hydrolysis for ARD1 but not ARF6"
 /db_xref="CDD:pfam01412"
 2052..2312
 /gene="DDEF2"
 /note="ANK; Region: ankyrin repeats"

misc_feature

/db_xref="CDD:cd00204"
 2991..3170
 /gene="DDEF2"
 /note="SH3; Region: Src homology 3 domains"
 /db_xref="CDD:smart00326"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 5544
 Score: 525.00 Matches: 790
 Percent Similarity: 99.00% Conservative: 0
 Best Local Similarity: 99.00% Mismatches: 4
 Query Match: 52.19% Indels: 8
 Gaps: 0
 US-09-914-042-1 (1-1006) x BC063308 (1-5544)
 Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
 Db 294 ATGCCGGACCAGATCTCCGTGTCGGAATTCGTGGCCGAGACCCATGAGGACTTCAAGGCG 353
 Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 Db 354 CCCACGGCCTCCAGCTTCCACCACCCGACGGCGCAGTCCCGGAAACACTGTGCGCGCCATC 413
 Qy 41 GluGluAlaLeuAspValAspAtgMetValLeuTyrLysMetLysLysSerValLysAla 60
 Db 414 GAGGAGGCTTTGGACGTGGACCGGATGTTCTTTCAAAATGAAGAAATCCGTGAAAGCA 473
 Qy 61 IleAsnSerSerGlyLeuAlaHisValIgluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
 Db 474 ATCAACAGCTCTGGCTGGCTCACGTGAAATGAAGACGAGTACACCCAGGCTCTGGAG 533
 Qy 81 LysPheGlyGlyAsnCys-ValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
 Db 534 AAGTTTGGCGCAACCG-TGTATGCAGAGATGACCCAGATTTAGGAAAGTGGTTCCTGAA 592
 Qy 100 sPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMetAs 120
 Db 593 GTTCTCAGTGTTTACAAGGAGGTTCACAGCACTTTTCAAAAACCTGTATTCAGATATGAA 652
 Qy 120 nAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysG 140
 Db 653 CAACATAATCTCTCCCTTTGGACAGTTTCTCTGAAGGGGACCTGAAAGGAGTGAAGG 712
 Qy 140 YAspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysI 160
 Db 713 CGATCTGAAAAGGCTTTTGTATAAGCTTGAAGGAGCTATGAAACAAATAAACCAAGAT 772
 Qy 160 eGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerG 180
 Db 773 AGAAAGAGAAAGAAAGAACACCGCAAGCTCCATGGGATGATTCGAGCTGAATAAGCGG 832
 Qy 180 YAlaGluLeuAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysG 200
 Db 833 AGCGAAATTCGCCAAGAGATGGAAGAGAGAGCGCTTCTTCCAGCTACAGATGTCGA 892
 Qy 200 uTyrLeuLeuLysValAsnGluLysLysIleLysLysGlyValAspLeuLeuGlnAsnLe 220
 Db 893 GTATCTGCTGAAGGTCAACGAAATCAAGATTAAAAAGGAGTAGATTACTTCAAGATCT 952
 Qy 220 uIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSe 240
 Db 953 GATCAATACCTTTTATGCCCAATGCAATTTTTTTTTCAGGATGAGTCAAGACCGTGAAG 1012
 Qy 240 rLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAs 260
 Db 1013 CCTCAACCTTCCATTGAAACGCTGTCTACGATCTTCAACGATCAACACAGGCCAGGA 1072
 Qy 260 pGluGluArgArgGlnLeuLysGlnLeuArgAspIleLeuLysSerAlaLeuGlnValG 280
 Db 1073 TCAAGAAAGAGGCGAGTTGTATACAGCTTCGAGATATTTTGAATATCCGATTCAGGTTGA 1132

Qy 280 uGlnlySGluAspSerGlnIleAtrGlnSerThrAlaTyrSerLeuHisGlnProGlnG1 300
Db 1133 ACAGAAAGAGAGCTCCCAAAATTCGTGAGACACAGCTTATAGCTTACATCAGCTCAGGG 1192
Qy 300 YAsnlySGluHisGlyThrGluAtrGlnSerLeuTyrLysLysSerAspGlyIleAr 320
Db 1193 AAACAGAGAAATGGGACCGAGCGGACGGAGCCTCTACAGAAAGATGACGGATCCG 1252
Qy 320 gLysValTrpGlnlyAsrGlySGysSerVallyAsnGlyPheLeuThrIleSerHisG1 340
Db 1253 AAAAGTGTGGCAGAAAGCAATGTTTCAGTTAAATAATGGTTTCTGACCATATCCCATGG 1312
Qy 340 YThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnVallyThrAsnPr 360
Db 1313 TACCGCTAACCGGCTCTCGAAAGCTCAACCTGCTAACTGCCAGGTGAAGCAACCC 1372
Qy 360 oGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaG1 380
Db 1373 TGAGGAGAAAGTGTCTTGACCTTATTTACATGACAGAACTTACCACTTTCAAGCTGA 1432
Qy 380 uAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLe 400
Db 1433 AGATGAACAGGAATGTCAAATATGATGTCTGTGCTGCAAAATAGCAAGAAAGACTTT 1492
Qy 400 uAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuTh 420
Db 1493 AAACAATGCATTTAAGGGGGATGACAATCTGAGAGAAATACATCGTCCAAAGAACTGAC 1552
Qy 420 rLysGluIleIleSerGluValGlnAtrGMetThrGlyAsnAspValCysCysAspCysG1 440
Db 1553 AAAGAGATCATCTCAGAAAGTGCAGAGATGACGGGCAATACGCTCTGTGACTGTGG 1612
Qy 440 YAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
Db 1613 GGGCCAGATCTACATGCTTTCCCAACCTGGCATCTCTGACCTGCATCGAGTGTTC 1672
Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspV 480
Db 1673 CGGAATCCACCGAGAGCTGGGGTTCATTATTCAG-GATGCAGTCCCTGACCTTAGATG 1731
Qy 480 alLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI 500
Db 1732 TACTGGGAACATCTGAGCTGTCTGCGCAAGAATATTTGGAAATGCGAGCTTTAATGAGA 1791
Qy 500 leMetGluCysCysLeuProAlaGluAspSerVallyAsnProGlySerAspMetA 520
Db 1792 TCATGGATGTGCTACCACTGAGGACTCAGTCNAACCCAAACCCAGCAGCAGCATGA 1851
Qy 520 snAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysH 540
Db 1852 ATGCAAGAAAGGACTTACATCAGCAAGTACATCGAGAGGAGATACGCAAGGAAGAAGC 1911
Qy 540 isAlaAspAsnAlaLysLeuHisSerLeuCysGluAlaVallyThrArgAspIleP 560
Db 1912 ACGCGGATAACCGCGGAGCTTACAGCTCTTCGAGGCGCTCAAAACGAGAGATATTT 1971
Qy 560 heGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaA 580
Db 1972 TTGGATGTCTCAGCTTATCTGATGTGTGGATCTTACGAAATAATCCCATGGCCA 2031
Qy 580 snGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrS 600
Db 2032 ACGGACATGAGCGGATGAAACGGCCCTCCACCTTGCACTGAGTCCGATCGATCGAACCT 2091
Qy 600 erLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyL 620
Db 2092 CTCTTCATGTAGACTTTTAGTTACAGACAGTGGGAACCTGGATTAACACAGACAGGA 2151
Qy 620 YGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuL 640
Db 2152 AAGGACGACAGCCCTGCACTACTGTGCTGACCGACATGCCAGTGCCTCAAGTTGC 2211
Qy 640 euLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspI 660

Db 2212 TCCTGCGGGGAAAGCCCTCCATCGAGATAGCAAAACGATGACGAGAGACTCCGCTGGACA 2271
Qy 660 leAlaLysAtrGLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyA 680
Db 2272 TTGCACAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGCTGACCCCAAGCCCTTATCGAA 2331
Qy 680 rgPheAsnSerHisValHisValGluTyrGluTyrTrpArgLeuLeuHisGluAspLeuAspG 700
Db 2332 GATTAAATTCACGTTTCAGCTTCAATATGATGAATGGCGACTACTCCACGAGACCTGGATG 2391
Qy 700 luSerAspAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArg 719
Db 2392 AAAGTGATGACGATCGATGAGAAATTTGACGCCAGTCC-CAACCGCGGGAAGACCG 2450
Qy 720 ProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAla 739
Db 2451 CCCATCAGCTTCTACAGCTGGGCTCCACAGCTTCTAGTCTAAGCTGTATCTTTGGCC 2510
Qy 740 ArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPheMetProSerIleLeuG1 759
Db 2511 AGAGATGCTGCAAACTTGCACAGGA-CAAGCAGAGGGCTTTTCATGCCAGCATCTTGCA 2569
Qy 759 nAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaPr 779
Db 2570 GAATGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGCAGCCCC 2629
Qy 779 oSerThrThrSerAlaProLeuProLeuProProArgAsnValGlyLys 794
Db 2630 CAGCACCCAGGCCCGCCCGCTTCTTCCACGGAATGTTGGCAA 2675

RESULT 3
174317
LOCUS I74317 4486 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 15 from patent US 5688641.
ACCESSION I74317
VERSION I74317.1 GI:3010458
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 4486)
AUTHORS Sager R., Zou, Z., Lee, S. Whan, and Tomasetto, C. Laure.
TITLE Cancer diagnosis using nucleic acid hybridization
JOURNAL Patent: US 5688641-A 15 18-NOV-1997;
FEATURES Location/Qualifiers
source 1..4486
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2,11e-292 Length: 4486
Score: 297.00 Matches: 701
Percent Similarity: 98.59% Conservative: 0
Best Local Similarity: 98.59% Mismatches: 4
Query Match: 29.52% Indels: 10
DB: Gaps: 0

US-09-914-042-1 (1-1006) x I74317 (1-4486)

Qy 302 LysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLys 321
Db 1 AAGGAACATGGGACCGAGCGGACGGCAGCCTCTACAAGAAGAGTGCAGGATCCGAAA 60
Qy 322 ValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341
Db 61 GTGTGGCAGAAAAGAAATGTTTCAATTTAAAAATGTTTCTGACCATATCCATGTTACC 120
Qy 342 AlaAsnArg-ProProAlaLysLeuAsnLeuLeuThrCysGlnVallyThrAsnProG1 361
Db 121 GCTAACCGGGGCTCTCTCAAGCTCAACCTGCTTAACCTGCCAGGTGAAGACCAACCTGA 180

Qy 361 uGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAs 381
Db 181 GGAGNAGAGTGTCTTGACCTTATTTTACATGACAGAACTTACCACCTTTCAAGCTGAAGA 240
Qy 381 pGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAs 401
Db 241 TGAACAGAAATGTCAAATATGATGCTGTGTCGCAAAATAGCAAAAGAAAGCTTTAA 300
Qy 401 nAsnAlaPheLysGlyAspAspAsnThrGly³HisAsnIleValGlnGluLeuThrLys 421
Db 301 CAATGCAATTAAGGGGGATGACAACTACCTGGAGAAATTAACATCGTCCAAAGAACTGACAAA 360
Qy 421 sGluIleIleSerGluValGlnAtqMetThrGlyAsnAspValCysCysAspCysGlyAl 441
Db 361 GGAGATCATCTCAGNAGTGCAGAGATGACGGGCAATGACGTCTGCTGTGACTGTGGGGC 420
Qy 441 aProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGl 461
Db 421 GCCAGATCCTACATGGCTTTCCACCAACTGGGCATCCTGACCTGCATCGAGTGTTCGG 480
Qy 461 yIleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerLeuThrLeuAspValL 481
Db 481 AATCCACGAGAGCTGGGGGTTCATTAATTCAG-GATCAGCTCCCTGACCTTAGATGTAC 539
Qy 481 euGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleM 501
Db 540 TGGGNACATCTGAGCTGTCTCGCAAGAATATTGGGAATGCAGGCTTTAATGAGATCA 599
Qy 501 etGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnA 521
Db 600 TGGAAATGTGCTACCACTGAGGACTCAGTCAAAACCCCAACCCAGGCGACATGAATG 659
Qy 521 laArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisA 541
Db 660 CAAGAAAGGACTACATCACCAAGTACATCGAGAGAGATACGCAAGGAGAGACG 719
Qy 541 laAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheG 561
Db 720 CGGATTAACGGCGGAGCTTCACAGTCTTTGGAGGCGCTCAAAACGAGAGATATTTTG 779
Qy 561 lyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnG 581
Db 780 GATTGCTCAAGCTTATGCTGATGTTGGATCTTACGGAAAAAATCCACTGGCCAAACG 839
Qy 581 lyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerL 601
Db 840 GACATGACCGCGATGAACGGCCCTCCACCTTGCACTGCACTGCGATCGGATCGAACCTCTC 899
Qy 601 euHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysG 621
Db 900 TTCACATTTGATAGCTTTTATGTTCAAGACAGTGGGAACTTGATTAACACAGACAGGAAAG 959
Qy 621 lySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuL 641
Db 960 GCAGCACAGCCCTGCCTACTGCTGCTGACCGCAATGCCAGTGCCTCAAGTTGCTCC 1019
Qy 641 euArgGlyValAsaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleA 661
Db 1020 TGGCGGGGAGGCGCTCCATCGAGATAGCAAAATGAGTCAAGGAGAGACTCCGCTGGCAATG 1079
Qy 661 laLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgP 681
Db 1080 CCAAGCGCTCAAGCACAGGACACTGTGAGGAGCTGCTGACCCAGACCTTATCTGGAGAT 1139
Qy 681 heAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGluS 701
Db 1140 TTAATCTCAGCTTCACTTGAATATGAATGCGGACTACTCCACGAAGACCTGGATGAA 1199
Qy 701 erAspAspMetAspGluLysLeuGlnProSerGlu-AsnArgArgGluAspArgPro 720
Db 1200 GTGATGACGATGATGAGAAATGACGCCAGCTCC-CAACCGCGGGAGACCGGCC 1258
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740

Db 1259 ATCAGCTTCTACAGCTGGGCTCCACAGAGCTTCACTCAACGCTGTATCTTTGGCCAGA 1318
Qy 741 AspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPheMetProSerIleLeuGlnAs 760
Db 1319 GATGCTCAAACTTGCCAAGGA-CAAGCAGAGGGCTTTTCATGCCAGCATCTTTGCAGAA 1377
Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSe 780
Db 1378 TGAGACTTTAGCGAGCCCTCTCTAGTGGAGCCCACTCCCGCCAGGCTGCGAGCCCCAG 1437
Qy 780 rThrThrSerAla-ProProLeuProProArgAsnValGlyLysValGlnThrAlaSerS 800
Db 1438 CACCACAGCGCCCGCCCGCTTCTCCACGGAATGTTGGCAAGATTTCAGACACCTCTCT 1497
Qy 800 erAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgS 820
Db 1498 CTGCTAAACACCTTGGAAAGACAACTCTGTAAGTGGAGCGTGGAGCGCGGAGCGAT 1557
Qy 820 erSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrA 840
Db 1558 CTTGCTCAGATCCCGCAGCTGTCCATCCACCGCTGCCCGCTTCTCGGCTGACATCTACCA 1617
Qy 840 snProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaL 860
Db 1618 ATCCCTTGACCCACCGCCCGCCCGCTTGGCAAGACGCCCGAGGTAATGAAGCCT 1677
Qy 860 euSerGlnProSerLysPheAlaProProGlyIleSerGlnIleArgProProProLeuP 880
Db 1678 TGAGCCAGCGGAGCAAGCTTGGCCGCTGGGATCTCACAGATCAGGCGCCCGCTCTGC 1737
Qy 880 roProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSer 899
Db 1738 CCCCACAGCGCCCGCCCGCTTCCCGCAGAAAGCTTGGCCAGGGGCT-GACAGTCC 1796
Qy 900 ThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAla 919
Db 1797 ACCCACTGACCAACAAGGCCAACCCGAGAGGACTTGTGGATCTCTCTGCAACCGAAGCT 1856
Qy 920 LeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSer 939
Db 1857 CTGGGCTCTCTGTCCAATGCTATGCTCTGAGCGCCCTGCAGCCCATGCGCTAGGAAGTCG 1916
Qy 940 GlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsn 959
Db 1917 CAGCAACCAAGTTGAAGCCTTAAGCGGTGAAGCGCTCTATAACTGTGTGGCTTGACAA 1976
Qy 960 ProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGln 979
Db 1977 CCGCATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGGACGGGAGGAGGACAG 2036
Qy 980 GluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSer 999
Db 2037 GAGTGTGGATTTGGCCACATTTGATGGAGATCTCTGTGCGAAAGCGCATTTCCCGGTGTCA 2096
Qy 1000 PheValHisPheIleAlaAsp 1006
Db 2097 TTTGTGCACCTTTATCGCTGAC 2117
RESULT 4
LOCUS I74314 4328 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 12 from patent US 5688641.
ACCESSION I74314
VERSION I74314.1 GI:3010455
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 4328)
AUTHORS Sager R., Zou, Z., Lee, S. Whan, and Tomasetto, C. Laure.
TITLE Cancer diagnosis using nucleic acid hybridization
JOURNAL Patent: US 5688641-A 12 18-NOV-1997;

FEATURES		Location/Qualifiers	
source		1..4328	
		/organism="unknown"	
ORIGIN		/mol_type="unassigned DNA"	
Alignment Scores:			
Pred. No.:	1..556-251	Length:	4328
Score:	257.00	Matches:	648
Percent Similarity:	98.63%	Conservative:	0
Best Local Similarity:	98.63%	Mismatches:	4
Query Match:	25.55%	Indels:	9
DB:	6	Gaps:	0
US-09-914-042-1 (1-1006) x 174314 (1-4328)			
QY	355	GlnValYthrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr	374
DB	3	CAGGTGAAGACCAACCTTGAGGAGAGAGTGTCTTGACCTTATTTTCATGACAGAACT	62
QY	375	TyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrrMetSerValLeuGlnAsn	394
DB	63	TACCACCTTTCAGCTGAAGTGAACAGGAAATGTCAAATATGATGCTGTGCTGCAAAAT	122
QY	395	SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn	414
DB	123	AGCAAGAGAGAGCTTTAAACAATGCATTTAAGGGGGATGACAATACTCGAGAGAAATAAC	182
QY	415	IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAsp	434
DB	183	ATCGTCCAGAACTGACAAAGAGAGATCATCTCAGAAAGTGACAGAGATGACGGGCAATGAC	242
QY	435	ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu	454
DB	243	GTCTGTGTGACTGTGGGGCGCCAGATCTACATGGCTTTCACCAACTTGGGCATCTCTG	302
QY	455	ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerPro-MetG1	474
DB	303	ACCTGCATCGAGTGTTCGGGAATCACCAGAGAGCTGGGGTTCATATTTCAG-GATGCA	361
QY	474	nSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAs	494
DB	362	GTCCCTGACCTTAGATGACTGGGAACATCTGAGCTGCTGCTCGCCCAAGAATATTGGGA	421
QY	494	nAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAs	514
DB	422	TGCAGCTTTAATGAGATCATGGAATGTTCCTACACAGCTGAGGACTCAGTCAAAACCA	481
QY	514	nProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgAr	534
DB	482	CCCAGGCAGCAGCATGAATGCAAGAAAGGACTATACATCAGCCAAAGTACATCGAGAGGAG	541
QY	534	GlyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuLeuValAla	554
DB	542	ATACCGAAGGAAGAGACGGCGGATTAACGGCGGAGGCTTCACAGCTTTTTCGAGAGCCGT	601
QY	554	IlysthrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrG1	574
DB	602	CAAAACGAGAGATATTTTTGGATGCTCCAGCTTATGCTGATGCTGATGCTGATGCTGATG	661
QY	574	uLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr	594
DB	662	AAAAATCCCACTGGCCCAACGAGACATGAGCCGATGAACGGCCCTCCACCTTGCAGTCA	721
QY	594	gSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLe	614
DB	722	ATCCGTGGGATCGAACCCTCTTTCACATGTTAGACTTTTTTAGTTTCAGAACAGTGGGAA	781
QY	614	uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAl	634
DB	782	GGATAAACAGACAGGGAAGGACAGCCCTGCACTACTGCTGCTGCTGCTGCTGCTGCTG	841
QY	634	aGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerG1	654

DB	842	CGAGTGCCTCAAGTTGCTTCTCGCGGGGAAGCCCTCCATCGAGATAGCAAAATAGACTCAGG	901
QY	654	yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlyGluLeuLeuTh	674
DB	902	AGAGACTCCGTGGACATTTGCCAAGCGCTTCAGCACGAGCAGCTGTGAGGAGCTCTGAC	961
QY	674	rGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrrArgLeuLe	694
DB	962	CCAAGCCTTATCTCGAAGATTAAATCTTCACGTTTCACTTCAATATGAATGCGGACTACT	1021
QY	694	uHisGluAspLeuAspGluSerAspAspAspMetAspGluLysLeuGlnProSerGlu-A	714
DB	1022	CCACGAAGACCTGGATGAAGTGAATGACGACATGGAATTCAGAAATTCAGAGCCAGTCC-CA	1080
QY	714	snArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerA	734
DB	1081	ACCGGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCACTCTA	1140
QY	734	snAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGlu-LysGlnArgAlaPhe	753
DB	1141	ACGCTGTATCTTTGGCCAGAGATGCTGCAAACTTGCACAGGA-CAAGCAGAGGGCTTTC	1199
QY	754	MetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProPro	773
DB	1200	ATGCCAGCATCTTGCAGAAATGAGACTTACGAGGCCCTCTGAGTGGCAGCCACCTCCC	1259
QY	774	AlaGlnProAlaAlaProSerThrSerAla-ProProLeuProProArgAsnValG1	793
DB	1260	CCCCAGCTGACGCCCCAGCACCCAGCCGCCGCCCTTCTCCACGGAAATGTTGG	1319
QY	793	yLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAs	813
DB	1320	CAAAATTCAGACAGCTCTCTCTGCTTAACACCTTGTGAAGACAAACTCTGTAAGTGTGA	1379
QY	813	pGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProPr	833
DB	1380	CGGTGGAGCGCGGAGCGATCTTCTGTCAGATCCGCCAGCTGTCTCCATCCACGCTGCC	1439
QY	833	oLeuArgValThrSerThrAsnProLeuThrProThrProProProValAlaLysTh	853
DB	1440	TCCTTGGCTGACATCTTACCAATCCCTGACCCCGCCGCGCCCGCCCGCTTGCACAGAC	1499
QY	853	rProSerValMetGluAlaLeuSerGlnProSerLysProAlaProGlyIleSerG1	873
DB	1500	GCCAGCGTAATGGAAGCTTTGAGCCAGCCGAGCAAGCTTCCCGCTGGGATCTCACA	1559
QY	873	nileArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAl	893
DB	1560	GATCAGGCCCCACCTCTGCCCCCAGCCGCCCGCCCTCCCGCAGAAAGAGCTGCTC	1619
QY	893	aPro-GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValA	913
DB	1620	GCCAGGGGCT-GACAAGTCCACCCCTGACCAAAAGGCGCAACCGAGAGACCTGTGG	1678
QY	913	spLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProA	933
DB	1679	ATCTCTGCAACGGAGCTCTGGGTCTCTCTGCTCAATGCTATGGTCTCGAGGCCCTCG	1738
QY	933	laProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuT	953
DB	1739	CACCATGCTTGGGAAGTCCGAGGCAACCAAGTTGAAGCCCTTAAGCGGTGAAAGCGCTCT	1798
QY	953	yrAsnCyValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleLev	973
DB	1799	ATAACTGTGTGGCTGACAAACCCCATGAGCTCACCTTCTCCGAGGGGATGTGATCATCG	1858
QY	973	alAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgL	993
DB	1859	TGGACGGGAGGAGGACGAGGATGGTGGATGGCCCATTTGATGGAGATCTCTGCTCGCA	1918
QY	993	yGlyAlaPheProValSerPheValHisPheIleAlaAsp	1006

Db	1919	AAGCGCATTC	CGGTGTCA	TTTGTGCAC	TTTATCGCTGAC	1959		
RESULT 5								
AKI72944			4713 bp	mRNA	linear	ROD 28-JUL-2004		
LOCUS	AKI72944							
DEFINITION	Mus musculus mRNA for mKIAA0400 protein.							
ACCESSION	AKI72944							
VERSION	AKI72944.1		GI:50510472					
KEYWORDS	FLI_CDNA.							
SOURCE	Mus musculus (house mouse)							
ORGANISM	Mus musculus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
AUTHORS	1 Okazaki, N., Kikuno, R.F., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S., Saga, Y., Seino, S., Nishimura, M., Kaisho, T., Hoshino, K., Kitamura, H., Nagase, T., Ohara, O. and Koga, H.							
TITLE	Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene: IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries							
JOURNAL	DNA Res. 11, 205-218 (2004)							
REFERENCE	2 (bases 1 to 4713)							
AUTHORS	Okazaki, N., Kikuno, R.F., Nagase, T., Ohara, O. and Koga, H.							
TITLE	Direct Submission							
JOURNAL	Submitted (19-MAY-2004) Hisashi Koga, Kazusa DNA Research Institute, Laboratory for Genome Informatics; 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan (E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918) The CREATE program supported by Japan science and technology corporation; cDNA full insert sequencing: Kazusa DNA Research Institute; cDNA library construction, clone selection and 5'- & 3'-end one pass sequencing.							
FEATURES	Location/Qualifiers							
source	1..4713							
	/organism="Mus musculus"							
	/mol_type="mRNA"							
	/db_xref="taxon:10090"							
	/clone="mfj04247"							
	/note="vector:modified pBC SK+"							
gene	1..4713							
	/gene="mKIAA0400"							
CDS	<65..2977							
	/gene="mKIAA0400"							
	/note="CDS is predicted by in silico analysis. Start codon is not identified."							
	/codon_start=1							
	/evidence="not experimental"							
	/product="mKIAA0400 protein"							
	/protein_id="BAD3222.1"							
	/db_xref="GI:50510473"							
	/translation="GGGSGPRPGVAMPDQISVSEFVAETHEDYKAPTASSFTTRTAQ CRNTVAIEEALDVRMVLVYRMKKSVKAINISGLAHVENEQQYTQALEKFGNCVCRD DPDLGSAFLKFSVFTKELTALFKNLIQNMNIIISPLDSLKLGLKGVKGLKXFPDK AWKDYETIKIEKEKHAHLGWMIRTEISGABIAEBEMEKERRFPQLOMCEYLLKVN EIKYKQDLTKLQNLKYHAQCNPQDGLKAVESLKPSIETSLDHLIKQAQDEER QLIQRDLKALQVEKESRDSQLRSTAYSLHQPOGNKEHGERNGNLYKKS DGI RKVMQKCSKWSGFLTISHGTANRPAPKLNLTQCVKTNPEEKCFPLISHDRTHF QAEDEBECQIWMVQLNSKEALNANAFKGDNDTGNNTVQELTKETISEVQRM TNDV CDCGAPDPTNLGLTICTEGSGIHELGVHYSRMQSLTLDLVLTSELLAKNIG NAGFNEIMECCPLSPBDVPKPNPNSDMIARKDYITAKYMERRYARKKHADTAALKHSLC EAVKTRDIFGLQAYADGVLTDEKIPLANGHEPDETALHLAVRSVDRSTSLHVDPLVQ NSGNDKDTGQSGSTALHYCCLTNNAECLKLLLRKASIEIANESGETPLDIAXRLKHE HCEBLLTQALSGRNSHVHYEYVWHLHEDLDESDDVDVEKLQSPNRRDRPVSYQ LGSSQFQSNVSLARDTANLTQKRGFGPSILQNETYGAILSGSPSSQIPFSTTS APPIPLPRNVGKDLTPPPPPVAKTSGTLEAMNOPSKSPGTSQSKPPPLPPPPSR LPQKPPASGTDKPTPLTNKGOPRGPPEASGPLSNMALQPPAPMPKSKQATSKPKRVK ALYNCVANPDELTFSEGDVITVDGEEDQEWIIGHIDGEPKRGAPFPVSVFHTAD"							
ORIGIN								
Alignment Scores:								
Pred. No.:	2.16e-153						4713	
Score:	161.00						Length: 4713 Matches: 694	

Percent Similarity:	95.99%	Conservative:	0
Best Local Similarity:	95.99%	Mismatch:	10
Query Match:	16.00%	Indels:	29
DB:	10	Gaps:	0
US-09-914-042-1 (1-1006) x AKI72944 (1-4713)			
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla	20
Db	101	ATGCCGGACCCAGATCTCCGTGTCGGAATTCGTGGCCGAGACCCATGAGGACTACAGCGC	160
Qy	21	ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaile	40
Db	161	CCACGCGCTTACCTCCACCGCGCACGCGCCAGTCCGCGAACACCGTGGCGCCATC	220
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla	60
Db	221	GAGGAGGCTTTGGACGTGGACCGGATGTTCTCTACAAAATGAAGAAATCTGTGAAGGCA	280
Qy	61	IleAsnSer-SerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuG1	80
Db	281	ATCAANTAT-CTCTGGCTGGCACACGTGGAGAACGAGAACGATACACCCAGGCTCTGGA	339
Qy	80	uLysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuL	100
Db	340	GAAGTTCGTTGGCAACTGTGTGTGCAGAGATGATCCAGACTTAGGAAGCGGTTCTGAA	399
Qy	100	sPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeGlnAsnMetAg	120
Db	400	GTTCTCTGTGTTACCAAGGAGTTGAGCGCACTCTTCAAAAACCTGATTCAGAACATGAA	459
Qy	120	nAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysG1	140
Db	460	CAACATCATCTCGTTCCTTGGACAGTTTCTGTAAGGAGATCTGAAGAGGATGAAGG	519
Qy	140	YAspLeuLysLysProPheAspLysAlaTPlLysAspTyrGluThrLysIleThrLysI1	160
Db	520	GGATCTGAAAAGGCCCTTTGATAAAGCTTGAAGGACTATGAAACAAAAATAACCAAAAT	579
Qy	160	eGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysIleSerG1	180
Db	580	AGAAAAGGAGAAAAAGGAGCATGCAAGCTCCACGGGATGATTCGTACTGAAATTAAGCGG	639
Qy	180	YAlaGluLleAlaGluMetGluLysGluArgPhePheGlnLeuGlnMetCysG1	200
Db	640	GCCTGAGATTCCAGGAGAGATGGAGAAAGAACGGTTCTTCCAGCTGCAGATGTGTGA	699
Qy	200	uTyrLeuLeuLysValAsnGluLysIle-LysLysGlyValAspLeuLeuGlnAsnL	220
Db	700	GTACCTGCTGAAGGTCAATGAAATCAAGGT-GAAAAAGGGAGTGGATTTGCTTCAGAAATC	758
Qy	220	eulLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluS	240
Db	759	TGATCAAGTACTTTTCAGCCCGAGTGCATTTTTCAGGATGGATTTGAAACAGCAGTAGAAA	818
Qy	240	erLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnA	260
Db	819	GCCTCAAGCCCTTCCATCGAGCGCTCTCCAGGACCTCCACCCATCAACAGGCCCCAGG	878
Qy	260	spGluGluArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValG	280
Db	879	ATGAGGAACCGCGACAGCTGATACAATTCAGATATTTTGAAGTCAGCATTCGAGGTGG	938
Qy	280	luGlnLysGlu-----AspSerGlnIle-ArgGlnSerThrAlaTyrSerLeuHis	296
Db	939	AACAGAAAGAGCTTAGGAGAGACTCACAACCT-TCGCCAAAGTACAGCCCTACAGCTTACAT	997
Qy	297	GlnProGlnGlyAsnLysGluHisGlyThrGluArgAsn-GlySerLeuTyrLysLysSe	316
Db	998	CAACCTCAGGGAACAAAGGAGCATGGAAACAGAGAGAACCGGAAC-CTCTACAGAGAGAG	1056
Qy	316	rAspGlyIleArgLysValTrpGlnLysArgLysSerValLysAsnGlyPheLeuTh	336

Db 1057 CGATGGGATCCGAAAGTGTGGCAGAGGAAGTGTTCCTTTAAAAATGCTTCTCTCAC 1116

Qy 336 rilesrhisglthrAlaAenArgProProAlaAlaLysLeuAenLeuThrCysGlnVa 356

Db 1117 CATCTCCACGGCACTGCGCAACCGGCCACCTGCGCAAGCTCAACCTGCTAACTGCCAGGT 1176

Qy 356 llystrAnProGluGluLysLysCysPheAspLeuLysSerHisAspArgThrTyRhi 376

Db 1177 GAAGACCAATCCCGAGGAGAAAGTGTTCGACCTCATATACATGATGACGAGCGTATCCA 1236

Qy 376 spheGlnAlaGluAenGluGlnGluCysGlnIleTirpMetSerValLeuGlnAenSerLys 396

Db 1237 CTTCAAGCGAGAGCGAGCAAGATGTCAGATATGGATGTCTGTACTGCGAGAACAGCA 1296

Qy 396 sGluGluAlaLeuAenAlaPheLysGlyAspAenThrGlyGluAenAenIleVa 416

Db 1297 GGAAGAAGCTCTGAACAACCGCTTTAAGGGTGTGACAACTGCGAGAAATATACATCGT 1356

Qy 416 lGlnGluLeuThrLysGluLysLysSerGluValGlnArgMetThrGlyAenAspValCy 436

Db 1357 CCAAGAGCTGACCAAGGAGATCATCTCGAGGTGCGAGGATGACGGCAATGACGTGTG 1416

Qy 436 sCysAspCysGlyAlaProAspProThrTirpLeuSerThrAenLeuGlyLleLeuThrCy 456

Db 1417 CTGGACTGTGGGGCACCAGATCCGACGTGCTCTTACCACCTGGGCATCTCTGACTTG 1476

Qy 456 sileGluCysSerGlyLleHisArgGluLeuGlyValHisTyrSerPro-MetGlnSerL 476

Db 1477 CATCGAGTCTCTGGATCCACCGGAGCTGGGGTTCATTACTCCAG-GATGAGTCCC 1535

Qy 476 euThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAenIleGlyAenAlaG 496

Db 1536 TGACATTAGATGTATTTGGGAACGTCTGAGCTTCTGCTTCCCAAGAAATATCGGAATGCG 1595

Qy 496 llyPheAenGluLleMetGluCysCysLeuProAla-GluAenSer-ValLysProAspPr 515

Db 1596 GCTTTAATGAATATATGAGGTGTTCCTTACCCTGTC-TGAGGACCC-AGTCAAGCCCCAACCC 1653

Qy 515 oGlySerAspMetAen-AlaArgLysAspTyrIleThrAlaLysTyrIle-GluArgArg 534

Db 1654 AGGAGTGCATGAT-TGCNAGGAGGACTACATCACAGCCAGTACAT-TGCAGAGGAGA 1711

Qy 535 TyrAlaArgLysLysHisAlaAspAen-AlaAlaLysLeuHisSerLeuCysGluAlaVa 554

Db 1712 TACGCACGGAAGCATGCGACAC-CGCGCGCAAGCTCCACAGCCCTTTCGAGCGCGT 1770

Qy 554 llystrArgAspIlePheGlyLeuGluAlaTyrAlaAspGlyValAspLeuThrG1 574

Db 1771 CAAGACCAGAGACATTTTGGGTACTCCAAGCTTATCTGTGTGGACCTGACAGA 1830

Qy 574 uLysIleProLeuAlaAenGlyHisGluProAspGluThrAlaLeuHisLeuAlaValAr 594

Db 1831 GAAATCCCATGCGCATGGCATGGCATGACAGATGAGACAGCCCTCCATCTTGCAGTCAG 1890

Qy 594 gserValAspArgThrSerLeuHisIleValAspPheLeuValGlnAenSerGlyAenLe 614

Db 1891 ATCTGTGACCGGACTTCCCTTCATTTAGACTTCTCGTCCAGAACAGTGGGAACCT 1950

Qy 614 uAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLysLeuThrAspAenAl 634

Db 1951 GGATAAACAGACAGCAGGCGGAGCAGACAGCCCTGCTGCTGCTGCGCAGACAAATGC 2010

Qy 634 aGluCysLeuLysLeuLeuAenGlyLysAlaSerIleGluLleAlaAenGluSerG1 654

Db 2011 TGAGTGCTCAAGCTCTCTCGGGGAAAGCCCTCAATTGATGATGATGATGATGATGATG 2070

Qy 654 yGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuTh 674

Db 2071 AGAGACCCCGTGGACATTTGCTAAGCGCTCAAGCAGCAACACTGTGAGGAACCTGCTCAC 2130

Qy 674 rGlnAlaLeuSerGlyArgPheAenSerHisValHisValGluTyrGluTirpArgLeuLe 694

Db 2131 TCAGGCTTGTGGGGCGGTTTAACTCCACGTTCCAGTTCGAGTATGAATGCGCGCTGCT 2190

Qy 694 uHisGluAspLeuAspGluSerAspAsp 704

Db 2191 GCATGAGGACCTGGATGAGAGCGACGACGAT 2221

RESULT 6

BC080847

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BC080847 5072 bp mRNA linear ROD 20-SEP-2004
Mus musculus gene model 592, (NCBI), mRNA (CDNA clone MGC:90837
IMAGE:6852788), complete cds.

BC080847

BC080847.1 GI:51593636

MGC.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 5072)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalios, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.,
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 5072)

Director MGC Project.

Direct Submission

Submitted (27-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>Contact: nisc.mgc@nih.gov

Akter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.L., Bouffard, G.G., Green, K., Brinkley, C.C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgoun, C., Vost, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, S.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: TRAK Plate: 176 Row: i Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

FEATURES		Location/Qualifiers	
source	1..5072	/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="CS7BL/6"	
		/db_xref="taxon:10090"	
		/clone="MGC:90837 IMAGE:6852788"	
		/tissue_type="Brain"	
		/clone_lib="NIH BMAP_G10"	
		/lab_host="DH10B"	
		/note="vector: pYX-ASC"	
	1..5072	/gene="Gm592"	
gene		/note="synonym: Ddef2"	
		/db_xref="LocusID:211914"	
		/db_xref="MGI:2685438"	
	289..2718		
		/gene="Gm592"	
		/codon_start=1	
		/product="gene model 592"	
		/protein_id="AAH80847.1"	
		/db_xref="GI:51593637"	
		/db_xref="LocusID:211914"	
CDS		/db_xref="MGI:2685438"	
		/translation="MPDOI SVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAALIEAL	
		VDRWLYKMKSVKAINISGLAHVENEEOYQALEKEGNCVCDDPDGLGSAFLKS	
		VFKELRALFKNLQNMNIIISFPDLSLLKGLDKGVKGDLLKPPDKAWKDYETKIKTI	
		EKEKEHAKLMGIRTEISGAIEAEEMEKERRFFQIQMCEYLLKVNIEIKVKGVDLLQ	
		NLIKYFHAQCFQDGLKAVESLPKSIETLSTDLHTIKQAQDEERRQLIQRDLKSA	
		LQVKEQSLQRSTAYSLHQPGNKEHGTENGLYKKSQGIKRWKRCXKNGF	
		LTIHGTLKNSLGAENFEIMECCLPSPDPVKNPQSDMIARKDYITAKYMERVARK	
		KHAPTAALKSLCEAVKTRDI.FGLLOAVADGVLDTEKIPLANGHEPDETHLAVRSV	
		DRTSLHIVFLVQNSGNDKQTKGSTALHYCCCLTDNAECLLLRKGASIEANESG	
ORIGIN		ETPLDLARKHECEELLTQALSGRFNSHVYEMRLLEDLDESDDVDKELQPS	
		PNRREDRVSYQSGSSQFNNAVSLARDTANLTKDKQRFQPSILQNETYGAISGS	
		PPSSQSI PPSTTAPPLPRNVGDKPLTTPPPVAKTSGTLEAMNQPSKSSQPGTSQ	
		SKPPLPPQPSRLPQKXPSGTDKPTPLTNKGOPRGPASGSPLSNMALOPPAPMPR	
		KSQATKSKPKVKALYNCVADNPDELITSEGDVIVDGEEDQEWIIGHIDSEPSRKA	
		FPVSFVHFIAI	
	Alignment Scores:		
	Pred. No.:	2,298-153	Length: 5072
	Score:	161.00	Matches: 337
	Percent Similarity:	97.68%	Conservative: 0
	Best Local Similarity:	97.68%	Mismatches: 4
	Query Match:	16.00%	Indels: 8
	DB:	10	Gaps: 0
	US-09-914-042-1 (1-1006) x BC080847 (1-5072)		
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleVala	20
	289	ATCGCGGACCAATCTCCGTGTCGGAATTCGTGGCCGAGACCATGAGGACTTACAAAGCG	348
	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaIle	40
	349	CCCACGGCCTTAGCTTACCACGCGCAGCGCCCATGTCGGAGAACACCGTGGCGGCATC	408
	41	GluGluAlaLeuAspValAspArgMetValLeuTyrIleValMetLysSerValIleVala	60
	409	GAGGAGGCTTGGACGTCGACCGGATGTTCTCTACAAATGAAGAAATCTGTGAAGCA	468
	61	IleAsnSer-SerGlyLeuAlaHisValGluAsnGluGluIntyrThrGlnAlaLeuG1	80
	469	ATCAATAT-CTCTGGCTGGCACAGTCGAGAACGAGAACAGTACACCCAGGCTCTGGA	527
	80	uIyPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuIy	100
	528	GAAGTTCGGTGGCAACTGTGTGAGAGATGATCCAGACTTAGGAAGCGCGTTCCTGAA	587
Qy	100	sPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAs	120
Db	588	GTTCTCTGTGTTCCACCAAGGAGTTGACGGCACTCTTCAAAAAACCTGATTCAACATGAA	647
	120	nAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIleG1	140
	648	CAACATCATCTCGTTCCCTTTGGACAGTTTCTGAAAGGAGATCTGAAAGGAGTAAAGG	707
	140	YAspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysI1	160
	708	GGATCTGAAAAGAGCCCTTTGNTAAAGCTTGAAGAGACTATGAACAAATAACCAAAAT	767
	160	eGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerG1	180
	768	AGAAAGAGAGAAAAGAGAGCATGCCAAGCTCCACGGGATGATTGCTACTGAAATAAGCGG	827
	180	YAlaGluIleAlaGluGluMetGluLysGluLysArgPhePheGlnLeuMetCysG1	200
	828	GGCTGAGATTGCAGAGGAGATGGAGAAGAAAGCGGTCTTCCAGCTGCGAGATGTGTGA	887
	200	uTyrLeuLeuLysValAsnGluIleLysIle-LysLysGlyValAspLeuLeuGlnAsnL	220
Db	888	GTACTCTGCTGAGGTCAATGAAATCAAGGT-GAAAAAGGGAGTGGATTGCTTCAAGATC	946
	220	eulleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluS	240
	947	TGATCAAGTACTTTTCCAGCCAGTCAATTTTCCAGGATGGATTGAAAGCAGTAGAAA	1006
	240	erLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnA	260
	1007	GCCTCAAGCCCTTCCATGAGACGCTCTCCAGGACCTCCACACCATCAACAGGCCAGG	1066
	260	spGluGluArgGlnLeuLeuArgAspIleLeuLysSerAlaLeuGlnValG	280
	1067	ATGAGGAACGCGACAGCTGATACACTTCGAGATATTTTGAAGTCAGCTGCGAGTGG	1126
	280	luGlnLysGluAspSerGlnIle-ArgGlnSerThrAlaTyrSerLeuHisGlnProGln	299
	1127	AACAGAAAGAGGACTCAACT-TCGCCAAGTACAGCCTACAGCTTACATCAACCTCAG	1185
	300	GlyAsnLysGluHisGlyThrGluArgAsn-GlySerLeuTyrLysLysSerAspGlyI1	319
Qy	1186	GGAAACCAAGGAGCATGGAACAGAGGAACCGGAAC-CTCTACAGAGAGCGGATGGAT	1244
	319	eArgLysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHi	339
	1245	CCGGAAGTGTGGCAGAGAGAGAGTGTTCGTTAAAAATGGCTTCCTCACCATCTCCA	1304
	339	sGlyThr 341	
	1305	CGGCACT 1311	
	RESULT 7		
	CQ720712	2109 bp	DNA
	LOCUS	Sequence 6646 from Patent WO02068579.	linear
	DEFINITION	CQ720712	
	ACCESSION	CQ720712	
	VERSION	CQ720712.1	GI:42281569
	KEYWORDS		
	SOURCE		
	ORGANISM	Homo sapiens (human)	
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1	Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.	
	AUTHORS		
	TITLE	Kits, such as nucleic acid arrays, comprising a majority of	
		human exons or transcripts, for detecting expression and other uses	
		thereof	
	JOURNAL	Patent: WO 02068579-A 6646 06-SEP-2002;	
		PE Corporation (NY) (US)	
	FEATURES	Location/Qualifiers	
	source	1..2109	
		/organism="Homo sapiens"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:9606"	

ORIGIN	US-09-914-042-1 (1-1006) x CQ720712 (1-2109)	DB:	US-09-914-042-1 (1-1006) x AX746789 (1-3123)	Gaps:	0
Alignment Scores:					
Pred. No.:	5.27e-122	Length:	2109		
Score:	130.00	Matches:	130		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	12.92%	Indels:	0		
DB:	6	Gaps:	0		
Qy	342 AlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnProGlu	361	Qy	841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu	860
Db	772 GCTAACCGGCTCTCTGCAAGCTCAACCTGCTAACTGCCAGGTGAAGACCAACCTGAG	831	Db	272 CCCTGACCCCGCCGCGCCCGCCCGCTTCCCAAGAGCGCCAGCGTAAATGAAGCCTTG	331
Qy	362 GluLysLysCysPheAspLeuLysHisAspArgThrTyHisPheGlnAlaGluAsp	381	Qy	861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro	880
Db	832 GAGAAAGAGTCTTGGCTTATTTTCATCATGACGAACTTACCACTTTTCAAGCTGAAGAT	891	Db	332 AGCCAGCGAGCAAGCTGCGCCGCTGGGATCTCACAGATCAGCGCCCGCCACCTCTGCC	391
Qy	382 GluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsn	401	Qy	881 ProGlnProProSerArgLeuProGlnLysLysProAlaPro-GlyThrAspLysSerTh	900
Db	892 GAACAGGAATGCTCAATATGATGCTGTCTGTCGCAAAATAGCAAGAGAGCTTTAAAC	951	Db	392 CCACAGCGCGCCAGCGCTCCCGCAGAGAAGCTGCGCGGGGGCT-GACAAGTCCAC	450
Qy	402 AsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys	421	Qy	900 rProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe	920
Db	952 AATGCATTTAAGGGGGATGACAACTACTCGAGAAATAATCACTCGTCCAGAACTGACAAG	1011	Db	451 CCCACTGACCAACAAAGCGCCCAACGAGAGGACCTGTGGATCTCTCTGCAACGGAGCTCT	510
Qy	422 GluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAla	441	Qy	920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerG1	940
Db	1012 GAGATCATCTCAGAAGTGCAGAGGATGACGGCGAATGACGCTCTGCTGATGCTGGGGCG	1071	Db	511 GGGTCTCTGCTCCCAATGCTATGCTGCTGCAGCGCCCTGCACCCCATGCTAGGAGTCCGA	570
Qy	442 ProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGly	461	Qy	940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyTrpAsnCysValAlaAspAsnPr	960
Db	1072 CCAGATCCTCATGGCTTTCCACCAACTGGGCATCTGACCTGCATCGAGTGTTCGGA	1131	Db	571 GGCAACCAAGTTGAAGCTTAAGCGGGTGAAGCGCTCTATAACTGTGTGGCTGACAACCC	630
Qy	462 IleHisArgGluLeuGlyValHisTySer 471		Qy	960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1	980
Db	1132 ATCCACCGAGAGCTGGGGGTTTCATTATTC 1161		Db	631 CGATGAGCTACCTTCTCCGAGGGGATGTGATCATCTGTGACGGGAGGAGGACACGA	690
RESULT 8			Qy	1000 eValHisPheIleAlaAsp 1006	
AX746789			Db	751 TGTGCACTTTATCGCTGAC 769	
LOCUS	AX746789	3123 bp			
DEFINITION	Sequence 314 from Patent EP1308459.	mRNA			
ACCESSION	AX746789	linear			
VERSION	AX746789.1	GI:32131177			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.				
TITLE	Full-length cDNA sequences				
JOURNAL	Patent: EP 1308459-A 314 07-MAY-2003;				
FEATURES	Helix Research Institute (JP) ; Research Association for Biotechnology (JP)				
source	Location/Qualifiers				
	1. 3123				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
ORIGIN					
Alignment Scores:					
Pred. No.:	2.02e-101	Length:	3123		
Score:	110.00	Matches:	165		
Percent Similarity:	98.80%	Conservative:	0		
Best Local Similarity:	98.80%	Mismatches:	1		
Query Match:	10.93%	Indels:	2		

Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizuehima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs

Nat. Genet. 36 (1), 40-45 (2004)

14702039

JOURNAL
PUBMED
REFERENCE
AUTHORS

Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T., Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R., Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, E., Momiyama, H., Onogawa, S., Kaeriyama, S., Satoh, N., Matsunawa, H., Takahashi, E., Kataoka, R., Kuga, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S., Terashima, Y., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project

Unpublished

3 (bases 1 to 3123)

Isogai, T. and Yamamoto, J.

JOURNAL
REFERENCE
AUTHORS
TITLE

Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan, cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
source

Location/Qualifiers
1..3123
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CTONG20000452"
/tissue_type="tongue, tumor tissue"
/clone_lib="CTONG2"
/note="Cloning vector: pME18SFL3"
146..772
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC03588.1"
/db_xref="GI:21749416"

CDS

/translation="MPSTLQNETYCALLSGSPPPAQAAPTSTAPPPLPPRVNGKPL
TPTTPPVAKTSPVMSALQSPKAPGISOIRPPPLPQPSPRLPQKPPAPGADKST
PLTNKGQPRGPVDLSATLGLSNAMVLQPPAMPKRSQATKLKPKRVKALYNVCVAD
NPDELTFSEGDVIIVDGEDQEWIGHIDGPRKGAPFVSVFIAD"

ORIGIN

Alignment Scores:

Pred. No.: 2,02e-101 Length: 3123
Score: 110.00 Matches: 165
Percent Similarity: 98.80% Conservative: 0
Best Local Similarity: 98.80% Mismatches: 1
Query Match: 10.93% Indels: 2
DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x AK091121 (1-3123)

Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Ddb 272 CCCCTGACCCCGCCGCGCCCGCCCGCTTCCCAAGACGCCCGCGTAATGGAAGCCTTG 331
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnLysArgProProProLeuPro 880
Ddb 332 AGCCAGCGGAGCAAGCTGCGCCCGCTGGGATCTCAGAGATCAGGCCCGCCACCTCTGCCC 391
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProProGlyThrLysSerTh 900
Ddb 392 CCACAGCGCGCCAGCGCTCCCGCAGAGAAGCTCGCGCGGGGCT-GACAAAGTCCAC 450
Qy 900 rProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920
Ddb 451 CCCACTGACCAACAAAGCCCAACCGAGAGGACCTGTGGATCTCTCTCAACGGAGGCTCT 510
Qy 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGl 940
Ddb 511 GGGTCCTCTGTCGAATGCTATGTCCTGCAGCCCGCTGCACCATGCTAGGAAGTCGCA 570
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyAsnCysValAlaAspAsnPr 960
Ddb 571 GGCACCAAGTTGAAGCTTAAGCGGGTGAAGGCGCTCTATAACTGTGTGGCTGACAAACC 630
Qy 960 oAspGluLeuThrPheSerGluGlyAspValLysLeuValAspGlyGluGluAspGlnGl 980
Ddb 631 CGATGAGCTCACCTTCTCCGAGGGGGATGTGATCATCTGTGACCGGGAGGAGGACCA 690
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Ddb 691 GTGGTGGATTGGCCACATTCATGAGATCTCTGGTGCAGAGGCGCAATCCCGGTGCATT 750
Qy 1000 eValHisPheIleAlaAsp 1006
Ddb 751 TGTGCACTTTATCGCTGAC 769
RESULT 10
AK124900
LOCUS
DEFINITION Homo sapiens cDNA FLJ42910 fis, clone BRHIP3021778, highly similar
to Homo sapiens development and differentiation enhancing factor 2
(DDEF2).
ACCESSION AK124900
VERSION AK124900.1 GI:34530818
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Euteleostomi;
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Euteleostomi;
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayaashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Ohayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,

Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamaashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039

2
Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S., Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hota, T., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watanabe, M., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEO human cDNA sequencing project
Unpublished
3 (bases 1 to 3870)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUN-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
Location/Qualifiers
1. 3870
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="BRHIP3021778"
/tissue_type="hippocampus"
/clone_lib="BRHIP3"
/note="cloning vector: pME18SFL3"

Alignment Scores:
Pred. No.: 2,42e-101 Length: 3870
Score: 110.00 Matches: 110
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.93% Indels: 0
DB: 9 Gaps: 0

US-09-914-042-1 (1-1006) x AK124900 (1-3870)

QY 897 AsplysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAla 916
DB 1187 GACAGTCCACCCCACTACCAACAAAGCCACCGAGAGGACCTGTGGATCTCTGCA 1246

QY 917 ThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetPro 936
DB 1247 ACGGAAGCTCTGGGTCTCTGTCCCAATGCTATGTCCTCGACGCCCTGCACCCATGCT 1306

QY 937 ArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuThrAsnCysVal 956
DB 1307 AGGAAGTCCGAGGCAACCAAGTTGAAGCCTTAAGCGGGTGAAGCGCTCTATACTGTG 1366

QY 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleileValAspGlyGlu 976
DB 1367 GCTGACACCCCGTAGCTACCTCTCCGAGGGGATGTGATCATCTGACCGGGAG 1426

QY 977 GluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996

DB 1427 GAGGACCAAGAGTGGTGGATTGGCCACATTGATGAGATCTCTGGTCGAAAGCGCATTC 1486

QY 997 ProValSerPheValHisPheIleAlaAsp 1006
DB 1487 CCGGTGTCATTGTGGCACTTTTATGCTGAC 1516

RESULT 11
CR385625 858 bp mRNA linear VRT 05-APR-2004
LOCUS
DEFINITION
Gallus gallus finished cDNA, clone CHEST386c18.
ACCESSION
CR385625
VERSION
CR385625.1 GI:46238384
KEYWORDS
SOURCE
Gallus gallus (chicken)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 858)
AUTHORS
Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E., Croning, M.D.R., Davies, R.M., Francis, M.D., Graham, D.V., Hubbard, S.J., Humphray, S.J., Hunt, P.J., Maddison, M., McLaren, S.R., Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G., Tickle, C. and Wilson, S.A.
Direct Submission
Submitted (05-APR-2004) Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: chicken@bma.umist.ac.uk
BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA sequencing project.
This sequence is from the
BBSCR/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection, from a library constructed by Elizabeth Bosch. cDNA was prepared from RNA extracted from heads, normalised, and poly A-trimmed.
EcoRI-NotI cut cDNA was then ligated into the vector. Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia coli DH10B.
Location/Qualifiers
1. 858
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisee"
/db_xref="taxon:9031"
/clone="CHEST386c18"
/clone_lib="CSEQCHN62"
/dev_stage="stage 36"

ORIGIN
Alignment Scores:
Pred. No.: 7,55e-100 Length: 858
Score: 108.00 Matches: 160
Percent Similarity: 98.77% Conservativity: 0
Best Local Similarity: 98.77% Mismatches: 1
Query Match: 10.74% Indels: 2
DB: 5 Gaps: 0

US-09-914-042-1 (1-1006) x CR385625 (1-858)

QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
DB 345 CAGAAAGAGGATTCAGATTTCGTGAGAGTACAGTTTACATTCAGCCTCAGGGA 404

QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
DB 405 AACAGGAAACATGCACCTGAACGAATGGAGTCTGTACAAGAAAGTGTGAATCAGA 464

QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
DB 465 AAAGTGTGGCAGAAAAGGAGTCTCAGTTTAAACGGTTTTCTTACAATTTCCCATGGT 524

QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnPro 360
DB 525 ACAGCTAACCGACCTCCAGCAAGCTCAATCTGTTAACTGCAAGTGAACAAACCA 584


```

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisArgThrTyrHisPheGlnAlaGlu 380
Db 585 GAGGAGAAAATAATGTTTGGCTCATATCAGATGACGAGATACCACTTCCAAGCAGAA 644
Qy 381 AspGluGlnGluCysGlnIleTrpMet-SerValLeuGlnAenSerLysGluGluAlaLe 400
Db 645 GATGACAGGAAATGTCATATATGGAC-ATCTGTTCTACAAAACAGCAGGAGGAAGCTTT 703
Qy 400 uAsnAenAlaPheLysGlyAspAsnThrGlyGluAenAenIleValGlnGluLeuTh 420
Db 704 AAATAATGCATTCAAAGGAGATGATACACAGGAGAGAAAATAATATTTGCCAGGAACGAC 763
Qy 420 rLysGluIleIleSerGluValGlnArgMetThrClyAenAspValCysCysAspCysG1 440
Db 764 AAAAGAGATTATATCTGAAGTCCAAAGATGACTGGAATGATGTGTGTGACTGTGG 823
Qy 440 YAla 441
Db 824 AGCA 827

RESULT 12
AC080162
LOCUS Homo sapiens BAC clone RP11-400L8 from 2, complete sequence.
AC080162 171653 bp DNA linear PRI 09-JAN-2002
AC080162 AC080162.7 GI:18093140
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171653)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 171653)
AUTHORS Cordes,M., Kozlowski,A., Dixon,R. and Boyer,E.
TITLE The sequence of Homo sapiens BAC clone RP11-400L8
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 171653)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 171653)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:15145269.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0400L08
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-385J23; the clone sequenced to the right is RP11-264H21. Actual start of this clone is at base position 1 of RP11-400L8; actual end is at base position 171653 of RP11-400L8.

Data from AC079782 was used to finish this clone, AC080162.

Sequence derived from a single plasmid subclone from base position 47301 to 47329.

FEATURES	source
Location/Qualifiers	1. 171653
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-400L8"
	/clone_lib="RPCI-11"
misc_feature	179..418
	/note="match to EST BG959460 (NID:gl14377631)"
repeat_region	346..627
	/rpt_family="Alu"
misc_feature	741..1147
	/note="match to EST BF929212 (NID:gl2327340)"
repeat_region	890..1053
	/rpt_family="Alu"
repeat_region	1539..1734
	/rpt_family="Alu"
repeat_region	1707..1734
	/rpt_family="(A)n"
repeat_region	2030..2054
	/rpt_family="AT-rich"
repeat_region	2079..2490
	/rpt_family="L1"
misc_feature	3105..3156
	/note="similar to Sus scrofa EST BG384893 (NID:gl3309365)"
misc_feature	3115..3156
	/note="match to EST C16709 (NID:gl571416)"
misc_feature	3120..3156
	/note="similar to Rattus norvegicus EST BF394200 (NID:gl1379064)"
misc_feature	3240..3406
	/note="match to EST C16709 (NID:gl571416)"
misc_feature	3240..3406
	/note="similar to Rattus norvegicus EST BF394200 (NID:gl1379064)"
misc_feature	3240..3406
	/note="similar to Sus scrofa EST BG384893 (NID:gl3309365)"
misc_feature	3312..3561
	/note="match to EST BF751012 (NID:gl2077688)"
misc_feature	3384..3406
	/note="similar to Homo sapiens EST BF804201 (NID:gl2133190)"

misc_feature 3731. 4236
/note="match to EST BF055560 (NID:g10809456)"
repeat_region 3805. 3897
/rpt_family="MIR"
misc_feature 4004. 4240
/note="match to EST AI708030 (NID:g4997806) as60b02.xl"
repeat_region 4258. 4557
/rpt_family="Alu"
misc_feature 5817. 5956
/note="similar to Rattus norvegicus EST BF394200 (NID:g11379064)"
misc_feature 5817. 5950
/note="match to EST AW996731 (NID:g8256965)"
misc_feature 5817. 5950
/note="similar to Homo sapiens EST BF804201 (NID:g12133190)"
misc_feature 5817. 5950
/note="similar to Sus scrofa EST BG384893 (NID:g13309365)"
repeat_region 5817. 5900
/note="match to EST C16709 (NID:g1571416)"
misc_feature 7266. 7491
/rpt_family="MIR"
7385. 7734
/note="match to EST AA377813 (NID:g2030152)"
repeat_region 7596. 7978
/rpt_family="MERI_type"
8578. 8648
/rpt_family="T-rich"
8620. 8924
/rpt_family="Alu"
misc_feature 8905
/note="match to EST H70683 (NID:g1042499) yul9f04.r1"
misc_feature 8942. 9396
/note="match to EST H70683 (NID:g1042499) yul9f04.r1"
repeat_region 9033. 9073
/rpt_family="(TG)n"
9454. 9481
/rpt_family="AT-rich"
9479. 9617
/rpt_family="L1"
9618. 9911
/rpt_family="Alu"
repeat_region 9912. 10645
/rpt_family="L1"
10646. 10924
/rpt_family="Alu"
11287. 11577
/rpt_family="Alu"
repeat_region 11870. 12043
/rpt_family="MIR"
14085. 14106
/rpt_family="AT-rich"
14089. 14392
/rpt_family="Alu"
14540. 14833
/rpt_family="Alu"
14936. 15102
/rpt_family="MIR"
15446. 15580
/note="match to EST H56609 (NID:g1005253) yc88d10.r1"
15486. 15580
/note="match to EST AW996731 (NID:g8256965)"
15486. 15580
/note="similar to Sus scrofa EST BG384893 (NID:g13309365)"
15486. 15566
/note="similar to Homo sapiens EST BF804201 (NID:g12133190)"
16075. 16092
/note="match to EST AI831610 (NID:g5452281) wj39g03.xl"
16079. 16378
/rpt_family="Alu"
16937. 17197
/rpt_family="Alu"

misc_feature 17082. 17601
/note="match to EST BF829427 (NID:g12174985)"
repeat_region 17585. 17753
/rpt_family="MIR"
Alignment Scores:
Pred. NO.: 2.53e-70 Length: 171653
Score: 81.00 Matches: 81
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.05% Indels: 0
DB: 9 Gaps: 0
US-09-914-042-1 (1-1006) x AC080162 (1-171653)
Qy 714 AenAArgGluAspArgProIleSerPheTyrrGlnLeuGlySerAsnGlnLeuGlnSer 733
|||||
Db 35364 AACCGGGGGGAGACCGGCCCATCAGCTTCTACAGCTGGGCTCAACACGCTTCAGTCT 35423
Qy 734 AenAlaValSerLeuAlaArgAspAlaAAsnLeuAlaLysGluLysGlnArgAlaPhe 753
|||||
Db 35424 AACGCTGTATCTTTGGCCAGAGATGCTGCAACCTTGCAAGAGAGACGAGGGCTTTC 35483
Qy 754 MetProSerIleLeuGlnAsnGluThrTyrrGlyAlaLeuLeuSerGlySerProProPro 773
|||||
Db 35484 ATGCCAGCATCTTTGCAGAAATGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCACCTCCC 35543
Qy 774 AlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGly 793
|||||
Db 35544 GCCAGGCTGACGCCCCAGCACCACCCAGGCCCGCTTCTCTCCACGAAATGTTGGC 35603
Qy 794 Lys 794
|||||
Db 35604 AAA 35606
RESULT 13
AKI23185
LOCUS
DEFINITION Homo sapiens cDNA FLJ41191 fis, clone BRACE2044946, weakly similar
to Homo sapiens ADP-ribosylation factor (arf)-directed GTPase
activating protein.
ACCESSION AKI23185
VERSION AKI23185.1 GI:34528666
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS
Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, E., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,
Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K.,
Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
Hirooka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
Yosida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,
Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,
Sasaki, N., Aota, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,
Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
Tashiro, H., Tanigami, A., Fujiiwara, T., Ono, T., Yamada, K., Fujii, Y.,
Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,
Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,

Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oiyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039
2

Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watanabe, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished

3 (bases 1 to 2157)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

LOCATION/Qualifiers
1..2157
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clones="BRACE2044946"
/tissue_type="cerebellum"
/clone_lib="BRACE2"
/note="Cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:
Pred. No.: 1.07e-61 Length: 2157
Score: 71.00 Matches: 71
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.06% Indels: 0
DB: 9 Gaps: 0
US-09-914-042-1 (1-1006) x AK123185 (1-2157)

Qy 372 AspArgThrTyrrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerVal 391
|
|
|
Db 1113 GACGAACTTACCCTTCAGCTGAAGATGAACAGGAATGTCAAAATGATGATGCTGTG 1172
|
|
|
Qy 392 LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGly 411
|
|
|
Db 1173 CTGCAAAATAGCAAGAGAGAGCTTTAAACAATGCTTTAAGGGGGATGACAACTACTGGA 1232
|
|
|
Qy 412 GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
|
|
|
Db 1233 GAAATAACATCGTCCAGAACTGCAAAAGGAGATCATCTCAGAGTCCAGAGGATGACG 1292
|
|
|
Qy 432 GlyAsnAspValCysCysAspCysGlyAlaPro 442
|
|
|
Db 1293 GGCAATGACCTGCTGTGACTGTGGGGGCCA 1325
|
|
|

RESULT 14
AR243296
LOCUS AR243296 2949 bp DNA linear PAT 20-DEC-2002

DEFINITION Sequence 8 from patent US 6475778.
ACCESSION AR243296
VERSION AR243296.1 GI:27290454
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2949)
AUTHORS Roberts, T.M., King, F.J., Harris, D.F., Hu, E., Spiegelman, B. and Chan, J.
TITLE Differentiation enhancing factors and uses therefor
JOURNAL Patent: US 6475778-A 8 05-NOV-2002;
FEATURES Location/Qualifiers
source 1..2949
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1.61e-58 Length: 2949
Score: 68.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.76% Indels: 0
DB: 6 Gaps: 0
US-09-914-042-1 (1-1006) x AR243296 (1-2949)

Qy 171 HisGlyMetIleArgThrGluIleSerGlyValaGluIleAlaGluMetGluLysGlu 190
|
|
|
Db 511 CACGGAATGATCCGACGGAGATCAGCGAGATAGCAGAGATAGCAGAGATAGCAGAGATAG 570
|
|
|
Qy 191 ArgArgPhePheGlnLeuGlnMetCysGluTyrrLeuLysValAsnGluIleLysIle 210
|
|
|
Db 571 CGCGCTTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAGTCAATGAATCAAGATC 630
|
|
|
Qy 211 LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrrPheHisAlaGlnCysAsnPhe 230
|
|
|
Db 631 AAAAAAGGTGCGACCTGCTCCAGAACTCATCAAACTATCTCCACGCACAGTGCAACTTC 690
|
|
|
Qy 231 PheGlnAspGlyLeuLysAlaVal 238
|
|
|
Db 691 TTTCCAGATGTCTCAAAAGCGGTG 714
|
|
|

RESULT 15

AR243295
LOCUS AR243295 5954 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 6 from patent US 6475778.
ACCESSION AR243295
VERSION AR243295.1 GI:27290453
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5954)
AUTHORS Roberts, T.M., King, F.J., Harris, D.F., Hu, E., Spiegelman, B. and Chan, J.
TITLE Differentiation enhancing factors and uses therefor
JOURNAL Patent: US 6475778-A 6 05-NOV-2002;
FEATURES Location/Qualifiers
source 1..5954
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.91e-58 Length: 5954
Score: 68.00 Matches: 68
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.76% Indels: 0
DB: 6 Gaps: 0
US-09-914-042-1 (1-1006) x AK123185 (1-2157)

Qy 372 AspArgThrTyrrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerVal 391
|
|
|
Db 1113 GACGAACTTACCCTTCAGCTGAAGATGAACAGGAATGTCAAAATGATGATGCTGTG 1172
|
|
|
Qy 392 LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGly 411
|
|
|
Db 1173 CTGCAAAATAGCAAGAGAGAGCTTTAAACAATGCTTTAAGGGGGATGACAACTACTGGA 1232
|
|
|
Qy 412 GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
|
|
|
Db 1233 GAAATAACATCGTCCAGAACTGCAAAAGGAGATCATCTCAGAGTCCAGAGGATGACG 1292
|
|
|
Qy 432 GlyAsnAspValCysCysAspCysGlyAlaPro 442
|
|
|
Db 1293 GGCAATGACCTGCTGTGACTGTGGGGGCCA 1325
|
|
|

RESULT 14

AR243296
LOCUS AR243296 2949 bp DNA linear PAT 20-DEC-2002

US-09-914-042-1 (1-1006) x AR243295 (1-5954)

Qy	171	HisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGlu	190
Db	943	CACGCAATGATCCGACGAGATCAGCGGACGAGATAGCAGAAAGAGATGAAAAAGAG	1002
Qy	191	ArgArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIle	210
Db	1003	CGGCGTTTCTTCCAGCTTCAGATGTGTGAGTACCTCTCAAAGTCAATGAAATCAAGATC	1062
Qy	211	LysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhe	230
Db	1063	AAAAAAGGTGCGACCTGCTCCAGAACTCATCAAAATCTTCCACGACACAGTGCAACTTC	1122
Qy	231	PheGlnAspGlyLeuLysAlaVal	238
Db	1123	TTTCAGGATGGTCTCAAGCGGTG	1146

Search completed: August 4, 2005, 22:16:00
Job time : 10766 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 00:25:21 ; Search time 113 Seconds
(without alignments)
4558.864 Million cell updates/sec

Title: US-09-914-042-1
Perfect score: 1006
Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGAFPVSFVHFIAID 1006

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1612378 seqs, 512079187 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	703	69.9	1006	1	DDF2_HUMAN	O43150 homo sapien
2	161	16.0	784	1	DDF2_MOUSE	Q7a196 mus musculus
3	161	16.0	809	2	Q66JN2	Q66jln2 mus musculus
4	161	16.0	970	2	Q6A074	Q6a074 mus musculus
5	110	10.9	208	2	Q8N282	Q8n282 homo sapien
6	46	4.6	1029	2	Q8NRK0	Q8nrk0 xenopus lae
7	34	3.4	108	2	Q8K115	Q8k115 mus musculus
8	30	3.0	1129	1	DDF1_BOVIN	O97902 b 130-kda p
9	30	3.0	1147	1	DDF1_MOUSE	Q9qwy8 m 130-kda p
10	28	2.8	111	2	Q8BYB8	Q8byb8 mus musculus
11	23	2.3	98	2	Q8K0U0	Q8k0u0 mus musculus
12	22	2.2	956	1	DDF1_HUMAN	Q9ulh1 h 130-kda p
13	19	1.9	123	2	Q8C1G6	Q8c1g6 mus musculus
14	16	1.6	121	2	Q8R1L5	Q8r1l5 mus musculus
15	15	1.5	155	2	Q9D308	Q9d308 mus musculus
16	14	1.4	104	2	Q9CYR1	Q9cyr1 mus musculus
17	14	1.4	112	2	Q8P562	Q8p562 mus musculus
18	14	1.4	192	2	Q8C3B6	Q8c3b6 mus musculus
19	12	1.2	1243	2	Q8TBN3	Q8tbn3 homo sapien
20	12	1.2	1243	2	Q6T7X3	Q6t7x3 homo sapien
21	12	1.2	1244	2	O00562	O00562 homo sapien
22	12	1.2	1244	2	Q9BZ73	Q9bz73 homo sapien
23	11	1.1	307	2	Q8DEH1	Q8deh1 brachydanio
24	10	1.0	118	2	Q8C0H9	Q8c0h9 mus musculus
25	10	1.0	138	1	YGJM_ECOL6	P67702 escherichia
26	10	1.0	138	1	YGJM_ECOLI	P67701 escherichia
27	10	1.0	138	1	YGJM_SHIFL	P67703 shigella fl
28	10	1.0	903	2	Q8TDY4	Q8tdy4 homo sapien
29	10	1.0	903	2	Q8P9F4	Q8p9f4 homo sapien
30	10	1.0	1358	2	Q7RWZ2	Q7rwz2 neurospora
31	10	1.0	1396	2	Q8WZY9	Q8wzy9 neurospora

32 9 0.9 125 2 Q8H4F8
33 9 0.9 202 2 Q8PJ46
34 9 0.9 281 2 Q95UG6
35 9 0.9 287 2 Q95UG5
36 9 0.9 309 2 Q9P6R1
37 9 0.9 322 2 Q8K0K1
38 9 0.9 356 2 Q9H3X2
39 9 0.9 360 2 Q96T14
40 9 0.9 366 2 Q811F3
41 9 0.9 403 2 Q86XV5
42 9 0.9 487 2 Q6GNT5
43 9 0.9 513 2 Q7RW49
44 9 0.9 519 2 Q6PA21
45 9 0.9 529 2 Q8BV81

Q8h4f8 oryza sativ
Q8pj46 homo sapien
Q95ug6 babesia bov
Q95ug5 babesia bov
Q9p6r1 schizosacch
Q8k0k1 mus musculu
Q9h3x2 homo sapien
Q96t14 homo sapien
Q811f3 mus musculu
Q86xv5 homo sapien
Q6gnt5 xenopus lae
Q7rw49 plasmodium
Q6pa21 xenopus lae
Q8bv81 mus musculu

ALIGNMENTS

RESULT 1
ID DDF2_HUMAN STANDARD; PRT; 1006 AA.
AC O43150;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Development and differentiation-enhancing factor 2 (Pyk2 C-terminus
DE associated protein) (PAP) (Paxillin-associated protein with ARFGAP
DE activity 3) (PAG3).
GN Name=DDEF2; Synonyms=KIAA0400;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=98116655; PubMed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VIII.
RT 78 new cDNA clones from brain which code for large proteins in
RT vitro.";
RL DNA Res. 4:307-313 (1997).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RP TISSUE=Placenta;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Hylk S.W.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Trimmwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinaki M.I., Skalak U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RL [3]
RP FUNCTION, PHOSPHORYLATION, INTERACTION WITH ARF1, ARF5, PTK2B
RP AND SRC, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND ALTERNATIVE
RP SPLICING.
RX PubMed=10022920;

RA Andreev J., Simon J.-P., Sabatini D.D., Kam J., Plowman G.,
 RA Randazzo P.A., Schlessinger J.;
 RT "Identification of a new Pyk2 target protein with Arf-GAP activity";
 RL Mol. Cell. Biol. 19:2338-2350(1999).
 [4]
 RP FUNCTION, MUTAGENESIS OF CYS-436, SUBCELLULAR LOCATION, AND
 RP INTERACTION WITH PXN.
 RX PubMed:1074932;
 RA Kondo A., Hashimoto S., Yano H., Nagayama K., Mazaki Y., Sabe H.;
 RT "A new paxillin-binding protein, PAG3/Papalpa/KIAA0400, bearing an
 RT ADP-ribosylation factor GTPase-activating protein activity, is
 RT involved in paxillin recruitment to focal adhesions and cell
 RT migration";
 RL Mol. Biol. Cell 11:1315-1327(2000).
 RN [5]
 RP FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH ARF6 AND ACTIN
 RP FILAMENTS.
 RX PubMed:11304556;
 RA Uchida H., Kondo A., Yoshimura Y., Mazaki Y., Sabe H.;
 RT "PAG3/Papalpa/KIAA0400, a GTPase-activating protein for ADP-
 RT ribosylation factor (ARF), regulates ARF6 in Fc gamma receptor-mediated
 RT phagocytosis of macrophages.";
 RL J. Exp. Med. 193:955-965(2001).
 CC -!- FUNCTION: Activates the small GTPases ARF1, ARF5 and ARF6.
 CC Regulates the formation of post-Golgi vesicles and modulates
 CC constitutive secretion. Modulates phagocytosis mediated by Fc
 CC gamma receptor and ARF6. Modulates FXN recruitment to focal
 CC contacts and cell migration.
 CC -!- SUBUNIT: Binds PXN, ARF1, ARF5, ARF6, PTK2B and SRC.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
 CC associated with the plasma membrane and with Golgi stacks.
 CC Colocalizes with F-actin and ARF6 in phagocytic cups.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=PAPalpha;
 CC IsoId=043150-1; Sequence=Displayed;
 CC Name=2; Synonyms=PAPbeta;
 CC IsoId=043150-2; Sequence=VSP_009722;
 CC -!- TISSUE SPECIFICITY: Detected in heart, brain, placenta, kidney,
 CC monocytes and pancreas.
 CC -!- INDUCTION: Up-regulated during monocyte maturation.
 CC -!- DOMAIN: The conserved Arg-464 in the Arf-GAP domain probably
 CC becomes part of the active site of bound small GTPases and is
 CC necessary for Gnp hydrolysis.
 CC -!- PTM: Phosphorylated on tyrosine residues by SRC and PTK2B.
 CC -!- SIMILARITY: Contains 2 ANK repeats.
 CC -!- SIMILARITY: Contains 1 Arf-GAP domain.
 CC -!- SIMILARITY: Contains 1 PH domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB007860; BAA23696.2; ALT_INIT.
 DR EMBL; BC063308; AAH63308.1; -.
 DR HSGP; O60631; IGBO.
 DR Genew; HGNC:2721; DDEF2.
 DR MIM; 603817; -.
 DR GO; GO:0008047; F:enzyme activator activity; TAS.
 DR ProDom; PD000066; SH3; 1.
 DR PROSITE; PS50297; ANK REP REGION; 1.
 DR PROSITE; PS50088; ANK REPEAT; 1.
 DR PROSITE; PS50115; ARFGAP; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW Alternative splicing; ANK repeat; Coiled coil; Golgi stack;
 KW GTPase activation; Metal-binding; Phosphorylation; Repeat; SH3 domain;
 KW Zinc.

FT	DOMAIN	305	397	PH.
FT	DOMAIN	421	543	Arf-GAP.
FT	REPEAT	584	616	ANK 1.
FT	REPEAT	620	649	ANK 2.
FT	DOMAIN	944	1006	SH3.
FT	DOMAIN	256	283	Coiled coil (Potential).
FT	DOMAIN	729	752	Coiled coil (Potential).
FT	DOMAIN	771	936	Pro-rich.
FT	VARSPPLIC	795	840	VQTASSANTLWKTNSVSDGSGRORSSDDPPAVHPPLPLR
FT				VTSTN -> D (in isoform 2).
FT				/FTId=VSP_009722.
FT	MUTAGEN	436	436	C -> R (in Ref. 2).
FT	CONFLICT	86	86	C -> R (in Ref. 2).
FT	CONFLICT	748	748	E -> D (in Ref. 2).
SQ	SEQUENCE	1006 AA;	111650 MW;	6A213517DCD995E1B CRC64;
	Query Match		69.9%;	Score 703; DB 1; Length 1006;
	Best Local Similarity		99.7%;	pred. No. 0;
	Matches 1003;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
Qy	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAATBEALDVRMVLKMKSKVKA	60	
Db	1	MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAATBEALDVRMVLKMKSKVKA	60	
Qy	61	INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGSFLKFSVFTKELTALFKNLIQNMN	120	
Db	61	INSSGLAHVENEEOYTQALEKFGNCVCRDDPDGLGSFLKFSVFTKELTALFKNLIQNMN	120	
Qy	121	NIISFPLDSLKGLKGVKGLKPKPKAMKDYETKITKEKEKEHAKLHGMIRTEISG	180	
Db	121	NIISFPLDSLKGLKGVKGLKPKPKAMKDYETKITKEKEKEHAKLHGMIRTEISG	180	
Qy	181	AEIAEEMEKERRPQLOWCEYLLKVNBIKKGVDLLQNLKIYFHAQCNFPQDGLKAVES	240	
Db	181	AEIAEEMEKERRPQLOWCEYLLKVNBIKKGVDLLQNLKIYFHAQCNFPQDGLKAVES	240	
Qy	241	LKPSIETLSTDLHTIKQAQDEERQQLQRLDILKSAQVEKESQSRQSTAYSLSHQPG	300	
Db	241	LKPSIETLSTDLHTIKQAQDEERQQLQRLDILKSAQVEKESQSRQSTAYSLSHQPG	300	
Qy	301	NKEHGTNRNGSLYKKSIGIRKVKQKCSVKNQGLTISHGTANRPPAKNLNLTQVKTNP	360	
Db	301	NKEHGTNRNGSLYKKSIGIRKVKQKCSVKNQGLTISHGTANRPPAKNLNLTQVKTNP	360	
Qy	361	BEKCFDLISHDRTHYFQAEDQEQCIWMSVLQNSKEALNNAFKGDDTENNIVQELT	420	
Db	361	BEKCFDLISHDRTHYFQAEDQEQCIWMSVLQNSKEALNNAFKGDDTENNIVQELT	420	
Qy	421	KEIISEVQRMGTGNDVCCGAPDPTWLSTNLGILTCIECSGIRHRELGVHYSQMOSLTLDV	480	
Db	421	KEIISEVQRMGTGNDVCCGAPDPTWLSTNLGILTCIECSGIRHRELGVHYSQMOSLTLDV	480	
Qy	481	LGTSELLAKNIGNAGNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYARKKH	540	
Db	481	LGTSELLAKNIGNAGNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYARKKH	540	
Qy	541	ADNAAKLHSLCEAVKTRDI FGLLQAYAGDVLTETKI PLANGHEDETHALHLAVSRVDRTS	600	
Db	541	ADNAAKLHSLCEAVKTRDI FGLLQAYAGDVLTETKI PLANGHEDETHALHLAVSRVDRTS	600	
Qy	601	LHIVDFLVQNSGNDLKOTGKGSTALHYCCLTDNAECLKLLLRGKASIEIANESGETPLDI	660	
Db	601	LHIVDFLVQNSGNDLKOTGKGSTALHYCCLTDNAECLKLLLRGKASIEIANESGETPLDI	660	
Qy	661	AKRLKHEHCEBELLTQALSGRPNSHVHVEYEWRLHEDLDESDDMDKELQSENNRDRP	720	
Db	661	AKRLKHEHCEBELLTQALSGRPNSHVHVEYEWRLHEDLDESDDMDKELQSENNRDRP	720	
Qy	721	ISFYQLGSNQLQSNVSLARDAAANLAKKORAFMPSILQNETYCALLSGSPPPQAPAPS	780	
Db	721	ISFYQLGSNQLQSNVSLARDAAANLAKKORAFMPSILQNETYCALLSGSPPPQAPAPS	780	
Qy	781	TTSAPPLPPRNVGKVVQTASSANTLWKTNSVSDGSGRORSSDDPPAVHPPLPLRVTSTN	840	


```

Db 781 TTSAPLPPRVKQVQTASSANTLWKTNSVVDGSGRSSDDPPAVPPPLPPPLRVSTN 840
Qy 841 PLTPPPPPVAKTPSVMSEALSQSPKAPPAGISQIRPPPLPQPSPRLPQKPKAPGTDKST 900
Db 841 PLTPPPPPVAKTPSVMSEALSQSPKAPPAGISQIRPPPLPQPSPRLPQKPKAPGADKST 900
Qy 901 PLTNKGQGRGVDSLSATGALGPLSNAMVLQPPAPMPKRSQATKLKPKRVKALYNCAVADNP 960
Db 901 PLTNKGQGRGVDSLSATGALGPLSNAMVLQPPAPMPKRSQATKLKPKRVKALYNCAVADNP 960
Qy 961 DELTFSEGDVLIIDGEEQDEWIGHIDGDPGRKGAPFVSFVHFAD 1006
Db 961 DELTFSEGDVLIIDGEEQDEWIGHIDGDPGRKGAPFVSFVHFAD 1006

RESULT 2
DDF2_MOUSE STANDARD; PRT; 784 AA.
AC Q7SIG6;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Development and differentiation-enhancing factor 2 (Pyk2 C-terminus
DE associated protein) (PAP) (Paxillin-associated protein with ARFGAP
DE activity 3) (PAG3).
GN Name=Bdef2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP CLONING, FUNCTION, PHOSPHORYLATION, INTERACTION WITH PTK2B AND SRC,
RP AND SUBCELLULAR LOCATION.
RX PubMed=10022920;
RA Andreev J., Simon J.-P., Sabatini D.D., Kam J., Plowman G.,
RA Randazzo P.A., Schllessinger J.;
RT "Identification of a new Pyk2 target protein with Arf-GAP activity.";
RL Mol. Cell. Biol. 19:2338-2350(1999).
RN [2]
RP FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH ARF6 AND ACTIN
RP FILAMENTS
RX PubMed=11304556;
RA Uchida H., Kondo A., Yoshimura Y., Mazaki Y., Sabe H.;
RA "PAG3/Papalpa/KIAA0400, a GTPase-activating protein for ADP-
RT ribosylation factor (ARF), regulates ARF6 in Pcgamma receptor-mediated
RT phagocytosis of macrophages.";
RL J. Exp. Med. 193:955-966(2001).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 245-522 IN COMPLEX WITH A
RP ZINC ION, AND MUTAGENESIS OF TRP-274; ILE-285; ARG-292; LEU-306 AND
RP ASP-307.
RX MEDLINE=20069319; PubMed=10601011; DOI=10.1093/emboj/18.24.6890;
RA Mandiyan V., Andreev J., Schllessinger J., Hubbard S.R.;
RT "Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-
RT associated protein beta.";
RL EMBO J. 18:6890-6898(1999).
CC -!- FUNCTION: Activates the small Gtpases ARF1, ARF5 and ARF6.
CC Regulates the formation of post-Golgi vesicles and modulates
CC constitutive secretion. Modulates phagocytosis mediated by Fc
CC gamma receptor and ARF6. Modulates PAXN recruitment to focal
CC contacts and cell migration (By similarity).
CC -!- SUBUNIT: Binds PAXN, ARF1, ARF5, ARF6, PTK2B and SRC.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC associated with the plasma membrane and with Golgi stacks.
CC Colocalizes with F-actin and ARF6 in phagocytic cups.
CC -!- DOMAIN: The conserved Arg-292 in the Arf-GAP domain probably
CC becomes part of the active site of bound small Gtpases and is
CC necessary for GTP hydrolysis.
CC -!- PTM: Phosphorylated on tyrosine residues by SRC and PTK2B (By
CC similarity).
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -!- SIMILARITY: Contains 1 Arf-GAP domain.

```

```

CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
DR PDB; IDQ; X-ray; A=1-278.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; hrIP_like.
DR Pfam; PF00023; Ank; 3.
DR Pfam; PF01412; ArfGAP; 1.
DR PRODOM; PD000066; SH3; 1.
DR PROSITE; PS50297; ANK REP REGION; 1.
DR PROSITE; PS50088; ANK REPEAT; 1.
DR PROSITE; PS50115; ArfGAP; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW 3D-structure; ANK repeat; Coiled coil; Goldi stack; GTPase activation;
KW Metal-binding; Phosphorylation; Repeat; SH3 domain; Zinc.
FT DOMAIN 84 111
FT COILED COIL (Potential).
FT DOMAIN 133 225
FT PH.
FT DOMAIN 249 371
FT ARF-GAP.
FT REPEAT 412 444
FT ANK 1.
FT REPEAT 448 477
FT ANK 2.
FT DOMAIN 599 714
FT PRO-rich.
FT DOMAIN 722 784
FT SH3.
FT MUTAGEN 274 274
FT W->A: Strongly reduces Arf-GAP mediated
FT stimulation of GTP hydrolysis.
FT I->A: Reduces Arf-GAP mediated
FT stimulation of GTP hydrolysis 100-fold
FT and abolishes Arf-GAP mediated
FT stimulation of GTP hydrolysis; when
FT associated with A-306.
FT R->K: Reduces Arf-GAP mediated
FT stimulation of GTP hydrolysis more than
FT 10000-fold.
FT R->A: Abolishes Arf-GAP mediated
FT stimulation of GTP hydrolysis.
FT L->A: Reduces Arf-GAP mediated
FT stimulation of GTP hydrolysis 100-fold
FT and abolishes Arf-GAP mediated
FT stimulation of GTP hydrolysis; when
FT associated with A-285.
FT D->A: Reduces Arf-GAP mediated
FT stimulation of GTP hydrolysis 1000-fold.
SQ SEQUENCE 784 AA; 87050 MW; F6775CFDA7870353 CRC64;

Query Match 16.0%; Score 161; DB 1; Length 784;
Best Local Similarity 100.0%; Pred.No. 4e-155;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPIANGHEPDETALHLAVRSVDRTSLHI 603
Db 372 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPIANGHEPDETALHLAVRSVDRTSLHI 431
Qy 604 VDFLVQNSGMLDKQTGKSTALHYCCLTDAECUKLLRGKASIEIANESGETPLDIANKR 663
Db 432 VDFLVQNSGMLDKQTGKSTALHYCCLTDAECUKLLRGKASIEIANESGETPLDIANKR 491
Qy 664 LKHEHCELLTQALSGRFNSHVHVEYEWRLIHEDLDESDDD 704
Db 492 LKHEHCELLTQALSGRFNSHVHVEYEWRLIHEDLDESDDD 532

RESULT 3
Q66JN2
ID Q66JN2 PRELIMINARY; PRT; 809 AA.
AC Q66JN2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Gene model 592.
GN Name=Gm592;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CS7BL/6; TISSUE=Brain;
RX  PubMed=12477932; DOI=10.1073/pnas.242603899;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA  Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA  Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA  Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA  Jones S.J., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length human
RT  and mouse cDNA sequences.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CS7BL/6; TISSUE=Brain;
RA  Director MGC Project;
RL  Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
CC  -1- SIMILARITY: Contains 1 SH3 domain.
DR  EMBL; BC080847; AAR80847.1; -.
DR  InterPro; IPR002110; ANK.
DR  InterPro; IPR000108; Neu_cyt_fact_2.
DR  InterPro; IPR011036; PH_related.
DR  InterPro; IPR001452; SH3.
DR  InterPro; IPR011511; SH3_2.
DR  Pfam; PF00023; Ank; 2.
DR  Pfam; PF00018; SH3_1; 1.
DR  Pfam; PF07653; SH3_2; 1.
DR  PRINTS; PR01415; ANKYRIN.
DR  PRINTS; PR00499; P67PHOX.
DR  ProDom; PD000066; SH3; 1.
DR  SMART; SM00248; ANK; 3.
DR  SMART; SM00326; SH3; 1.
DR  PROSITE; PS50088; ANK_REPEAT; 1.
DR  PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR  PROSITE; PS50002; SH3; 1.
DR  SMART; SM00248; ANK; 3.
DR  PROSITE; PS50088; ANK_REPEAT; 1.
DR  PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR  PROSITE; PS50002; SH3; 1.
DR  ANK repeat; SH3 domain.
KW  ANK repeat; SH3 domain.
SQ  SEQUENCE 809 AA; 90021 MW; B77505164F392CCA CRC64;

Query Match 16.0%; Score 161; DB 2; Length 809;
Best Local Similarity 100.0%; Pred. No. 4.1e-155;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTLSHI 603
Db 398 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTLSHI 457

Qy 604 VDFLVQNSGNLDKQTKGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 663
Db 458 VDFLVQNSGNLDKQTKGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 517

Qy 664 LKHEHCELLTQALSGRFSNVHVEYEWRLHEDLDESDDD 704
Db 518 LKHEHCELLTQALSGRFSNVHVEYEWRLHEDLDESDDD 558

RESULT 4
Q6A074 PRELIMINARY; PRT; 970 AA.
AC Q6A074;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

```

```

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE MKTAA0400 protein (Fragment).
GN Mus musculus (Mouse).
OS Eukaryota; Euteleostomi;
OC Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Okazaki N., Kikuno R.P., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Saga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-Fractionated Libraries.";
RL DNA Res. 11:205-218 (2004).
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AK172944; BA032222.1; -.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR011036; PH_related.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR011511; SH3_2.
DR Pfam; PF00023; Ank; 3.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00018; SH3_1; 1.
DR Pfam; PF07653; SH3_2; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00499; P67PHOX.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ArfGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
DR ANK repeat; SH3 domain.
KW ANK repeat; SH3 domain.
FT NON TER 1
SQ SEQUENCE 970 AA; 107768 MW; 11BAA16234DFEEC1 CRC64;

Query Match 16.0%; Score 161; DB 2; Length 970;
Best Local Similarity 100.0%; Pred. No. 4.8e-155;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTLSHI 603
Db 559 AAKLSLCEAVKTRDIFGLQAYADGVLDLTKIPLANGHEPDETALHLAVRSVDRSTLSHI 618

Qy 604 VDFLVQNSGNLDKQTKGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 663
Db 619 VDFLVQNSGNLDKQTKGKSTALHYCCCLTDNAECLKLLRGKASIEIANESGETPLDIAKR 678

Qy 664 LKHEHCELLTQALSGRFSNVHVEYEWRLHEDLDESDDD 704
Db 679 LKHEHCELLTQALSGRFSNVHVEYEWRLHEDLDESDDD 719

RESULT 5
Q6N282 PRELIMINARY; PRT; 208 AA.
AC Q6N282
ID Q6N282
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

```

DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ33802.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tongue;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamiyama K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuki H., Oshima A., Sasaki N., Rotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Moniyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsuura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
CC !- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; AK091121; BAC03588.1; -.
DR HSSP; Q9NZM3; 1UFF.
DR InterPro; IPR00108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3_1; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 208 AA; 21869 MW; 8ABA541403C9986F CRC64;
Query Match 10.9%; Score 110; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.2e-103;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 897 DKSTPLTKNGQPRGPDVLSATEALGPLSNAMVLPAPMPKRSQATKLKPKRVKALYNCV 956
DB 99 DKSTPLTKNGQPRGPDVLSATEALGPLSNAMVLPAPMPKRSQATKLKPKRVKALYNCV 158
QY 957 ADNPDELTFSEGDVLIIVDGEEDQEWIIGHIDGPRKGAFFVSVFHFAD 1006
DB 159 ADNPDELTFSEGDVLIIVDGEEDQEWIIGHIDGPRKGAFFVSVFHFAD 208
RESULT 6
QSNRKO PRELIMINARY; PRT; 1029 AA.
ID Q6NRKO
AC Q6NRKO;
DT 05-JUL-2004 (TReMBLrel. 27, Created)

DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE MG83760 protein.
GN Name=MG83760;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocytes;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC !- SIMILARITY: Contains 1 PH domain.
CC !- SIMILARITY: Contains 1 SH3 domain.
DR EMBL; BC070750; AAH70750.1; -.
DR HSSP; P19878; 1K4U.
DR GO; GO:0043087; P:regulation of GTPase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001164; ArfGAP.
DR InterPro; IPR001849; PH.
DR InterPro; IPR01036; PH related.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF01412; ArfGAP; 1.
DR Pfam; PF00169; PH; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00405; REVINTRACTNG.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00105; ArfGAP; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50115; ARFGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; SH3 domain.

DR InterPro; IPR001849; PH.
 DR InterPro; IPR001036; PH.related.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00023; Ank; 2.
 DR Pfam; PF01412; ArfGap; 1.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00499; P67PHOX.
 DR PRINTS; PR00405; REVINTRACNG.
 DR PRODOM; PD00066; SH3; 1.
 DR SMART; SM00248; Ank; 2.
 DR SMART; SM00105; ArfGap; 1.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50297; ANK REP REGION; 1.
 DR PROSITE; PS50088; ANK REPEAT; 2.
 DR PROSITE; PS50115; ARFGAP; 1.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW ANK repeat; Direct protein sequencing; GTPase activation; Membrane;
 FT Repeat; SH3 domain; Zinc-finger.
 FT DOMAIN 327 419 PH.
 FT DOMAIN 442 565 Arf-GAP.
 FT ZN FING 457 480 C4-type.
 FT REPEAT 603 635 ANK 1.
 FT REPEAT 639 668 ANK 2.
 FT DOMAIN 786 1059 Pro-rich.
 FT DOMAIN 1067 1129 SH3.
 FT CONFLICT 65 65 D -> T (in Ref. 2; AA sequence).
 SQ SEQUENCE 1129 AA; 125381 MW; C1576CEACQACDAB CRC64;
 Query Match 3.0%; Score 30; DB 1; Length 1129;
 Best Local Similarity 100.0%; Pred. No. 9.7e-21;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 127 LDSLLKGLKGVKGLKPFKPKWKDYETK 156
 |||||
 Db 147 LDSLLKGLKGVKGLKPFKPKWKDYETK 176
 RESULT 9
 DDFF1_MOUSE STANDARD; PRT; 1147 AA.
 AC Q9QWY8; O08612; Q80T68; Q80UV6; Q99LV8; Q922B6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE 130-kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-
 DE activating protein (PIP2-dependent ARF1 GAP) (ADP-ribosylation factor-
 DE directed GTPase-activating protein 1) (ARF GTPase-activating protein
 DE 1) (Development and differentiation-enhancing factor 1)
 DE (Differentiation-enhancing factor 1) (DEF-1).
 GN Name=Def1; Synonyms=Asap1, Kiaa1249, Shag1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), CHARACTERIZATION, AND
 RP MUTAGENESIS OF ARG-811; PRO-910 AND PRO-913.
 RC TISSUE=Brain, and Embryo;
 RX MEDLINE=99038209; PubMed=9819391;
 RA Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,
 RA Randazzo P.A.;
 RT "ASAP1, a phospholipid-dependent arf GTPase-activating protein that
 RT associates with and is phosphorylated by Src.";
 RL Mol. Cell. Biol. 18:7038-7051(1998).
 RN [2]
 RP SEQUENCE OF 57-1147 FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 RA Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT I1. The complete nucleotide sequences of 400 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries.";
 RL DNA Res. 10:35-48(2003).
 RN [3]
 RP SEQUENCE OF 654-1147 FROM N.A. (ISOFORM 1).
 RX MEDLINE=97271433; PubMed=9126384; DOI=10.1006/abio.1997.2040;
 RA Yamabhai M., Kay B.K.;
 RT "Examining the specificity of Src homology 3 domain -- ligand
 RT interactions with alkaline phosphatase fusion proteins.";
 RL Anal. Biochem. 247:143-151(1997).
 RN [4]
 RP SEQUENCE OF 20-1147 FROM N.A. (ISOFORM 4), AND SEQUENCE OF 722-1147
 RP FROM N.A. (ISOFORM 1).
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.2426603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Smerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=99147067; PubMed=10022919;
 RA King P.J., Hu E., Harris D.F., Sarraf P., Spiegelman B.M.,
 RA Roberts T.M.;
 RT "DEF-1, a novel src SH3 binding protein that promotes adipogenesis in
 RT fibroblastic cell lines.";
 RL Mol. Cell. Biol. 19:2330-2337(1999).
 CC -1- FUNCTION: May function as a signal transduction protein involved
 CC in the differentiation of fibroblasts into adipocytes and possibly
 CC other cell types (By similarity). Possesses phosphatidylinositol
 CC 4,5-bisphosphate-dependent GTPase-activating protein activity for
 CC ARF1 (ADP ribosylation factor 1) and ARF5 and a lesser activity
 CC towards ARF6. May coordinate membrane trafficking with cell growth
 CC or actin cytoskeleton remodeling by binding to both SRC and PIP2.
 CC -1- ENZYME REGULATION: Activity stimulated by phosphatidylinositol
 CC 4,5-bisphosphate (PIP2).
 CC -1- SUBUNIT: Homodimer. Interacts with SRC and CRK.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; predominantly. Membrane-
 CC associated; partially.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=4;
 CC Name=1; Synonyms=SHAG1a, ASAP1a;
 CC IsoId=Q9QWY8-1; Sequence=Display;
 CC Name=2; Synonyms=SHAG1b, ASAP1b;
 CC IsoId=Q9QWY8-2; Sequence=VSP_008368;
 CC Name=3;
 CC IsoId=Q9QWY8-3; Sequence=VSP_008366;
 CC Name=4;
 CC IsoId=Q9QWY8-4; Sequence=VSP_008367;
 CC -1- TISSUE SPECIFICITY: Expressed in all tissues examined but a most
 CC abundant expression was found in the testis, brain, lung and
 CC spleen. A heightened expression was seen in the adipose tissue
 CC from obese (ob) and diabetic (db) animals.
 CC -1- DOMAIN: The PH domain most probably contributes to the
 CC phosphoinositide-dependent regulation of ADP ribosylation factors.

CC -|- PTM: Phosphorylated on tyrosine residues by SRC.
 CC -|- SIMILARITY: Contains 2 ANK repeats.
 CC -|- SIMILARITY: Contains 1 Arf-GAP domain.
 CC -|- SIMILARITY: Contains 1 C4-type zinc finger.
 CC -|- SIMILARITY: Contains 1 PH domain.
 CC -|- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF075461; AAC98349.1; -;
 CC EMBL; AF075462; AAC98350.1; -;
 CC EMBL; AK122477; BAC65759.1; -;
 CC EMBL; BC002201; AAH02201.1; ALT_INIT.
 CC EMBL; BC048818; AAH48818.1; ALT_INIT.
 CC EMBL; U92478; AAB82338.1; -;
 CC PIR; T42627; T42627.
 CC HSSP; P15891; IJ08.
 CC MGD; MGI:1342335; Ddef1.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR001164; hRIP_like.
 CC InterPro; IPR00108; Neu_cyt_fact_2.
 CC InterPro; IPR001849; PH.
 CC InterPro; IPR011036; PH_related.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00023; Ank; 3.
 CC Pfam; PF01412; ArfGap; 1.
 CC Pfam; PF00169; PH; 1.
 CC Pfam; PF00018; SH3; 1.
 CC PRINTS; PR00499; P67PHOX.
 CC PRINTS; PR00405; REVINTRACTING.
 CC PRODOM; PD00066; SH3; 1.
 CC SMART; SM00248; ANK; 2.
 CC SMART; SM00105; ArfGap; 1.
 CC SMART; SM00233; PH; 1.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS0297; ANK_REPEAT; 1.
 CC PROSITE; PS0088; ANK_REPEAT; 2.
 CC PROSITE; PS0115; ARFGAP; 1.
 CC PROSITE; PS0003; PH_DOMAIN; 1.
 CC PROSITE; PS00002; SH3; 1.
 KW Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;
 KW SH3 domain; zinc-finger.
 FT DOMAIN 339 431 PH.
 FT DOMAIN 454 577 Arf-CAP.
 FT ZN_FING 469 492 C4-type.
 FT REPEAT 615 647 ANK 1.
 FT REPEAT 651 680 ANK 2.
 FT DOMAIN 798 1011 PRO-rich.
 FT DOMAIN 1085 1147 SH3.
 FT VARSPPLIC 304 315 Missing (in isoform 3).
 FT VARSPPLIC 304 318 Missing (in isoform 4).
 FT VARSPPLIC 304 318 Missing (in isoform 5).
 FT VARSPPLIC 816 872 Missing (in isoform 2).
 FT VARSPPLIC 816 872 Missing (in isoform 3).
 FT MUTAGEN 811 811 R->A: Significant reduction in binding to
 FT SRC and loss of phosphorylation.
 FT Loss of binding and phosphorylation; when
 FT associated with A-910 and A-913.
 FT P->A: Significant reduction in binding to
 FT SRC and CRK and decrease in
 FT phosphorylation; when associated with A-
 FT 913. Loss of binding and phosphorylation;
 FT when associated with A-811 and A-913.
 FT P->A: Significant reduction in binding to
 FT SRC and CRK and decrease in
 FT phosphorylation; when associated with A-
 FT 913.
 FT MUTAGEN 913 913

FT 910. Loss of binding and phosphorylation;
 FT when associated with A-811 and A-910.
 FT T->S (in Ref. 4).
 FT S->L (in Ref. 3 and 4).
 FT R->I (in Ref. 4).
 SQ SEQUENCE 1147 AA; 127395 MW; 1A08321C491B4609 CRC64;
 Query Match 3.0%; Score 30; DB 1; Length 1147;
 Best Local Similarity 100.0%; Pred. NO. 9.8e-21;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 127 LDSLLKGLKGVKGLKKPKPKAWKDYETK 156
 |||||
 Db 147 LDSLLKGLKGVKGLKKPKPKAWKDYETK 176
 |||||
 RESULT 10
 Q8BYB8 PRELIMINARY; PRT; 111 AA.
 ID Q8BYB8
 AC Q8BYB8
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-JUN-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
 DE library, clone:A630001111 product:K1AA0400 PROTEIN homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20499174; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
 RA Kono H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";

```

RL  Genome Res. 10:1757-1771(2000).
RN  [6]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57BL/6J; TISSUE=Thymus;
RA  Adachi J., Aizawa K., Akimura T., Hara A., Hashizume W.,
RA  Fukuda S., Furuno M., Hanagaki T., Hiramoto K., Hiraoka T., Hirozane T.,
RA  Hayashida K., Hayatsu N., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA  Horii F., Imotani K., Isihara Y., Kondo S., Konno H., Kouda M., Koya S.,
RA  Katoch H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA  Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA  Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA  Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA  Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA  Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA  Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RA  Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AK041326; BAC30907.1; -
SQ  SEQUENCE 111 AA; 11591 MW; 551D3D2E51983077 CRC64;

Query Match 2.8%; Score 28; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.4e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 MPDQISVSEFVAETHEDYKAPTASFTT 28
Db  1 MPDQISVSEFVAETHEDYKAPTASFTT 28

RESULT 11
Q8K0U0 PRELIMINARY; PRT; 98 AA.
ID  Q8K0U0
AC  Q8K0U0;
DT  01-OCT-2002 (TrEMBLrel. 22, Created)
DT  01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Similar to RIKEN CDNA 903062423 gene.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  SEQUENCE FROM N.A.
RC  TISSUE=Eye;
RX  MEDLINE=2328257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA  Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley D.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA  Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
RA  Jones S.J., Marra M.A.;
RA  "Generation and initial analysis of more than 15,000 full-length human
RT  and mouse cDNA sequences."
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Eye;
RA  Strausberg R.;
RA  Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; BC030401; AAH30401.1; -
DR  GO; GO:0005622; C:intracellular; IEA.
DR  GO; GO:0003676; F:nucleic acid binding; IEA.
DR  GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

```

```

DR  InterPro; IPR001909; KRAB.
DR  Pfam; PF01352; KRAB; 1.
DR  SMART; SM00349; KRAB; 1.
DR  PROSITE; PS08005; KRAB; 1.
SQ  SEQUENCE 98 AA; 10971 MW; 3BF7702A67E3535A CRC64;

Query Match 2.3%; Score 23; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.6e-14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  20 APTASSFTTRTAQCRNTVAALIEE 42
Db  20 APTASSFTTRTAQCRNTVAALIEE 42

RESULT 12
DDP1_HUMAN STANDARD; PRT; 956 AA.
ID  DDP1_HUMAN
AC  Q9ULH1;
DT  10-OCT-2003 (Rel. 42, Created)
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  05-JUL-2004 (Rel. 44, Last annotation update)
DE  130-kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-
DE  activating protein (PIP2-dependent ARF1 GAP) (ADP-ribosylation factor-
DE  directed GTPase-activating protein 1) (ARF GTPase-activating protein
DE  1) (Development and differentiation-enhancing factor 1) (Fragment).
GN  Name=DDP1; Synonyms=ASAP1, KIAA1249;
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  SEQUENCE FROM N.A. (ISOFORM 2).
RC  TISSUE=Bone marrow;
RA  Wambutt R., Heubner D., Mewes H.-W., Weil B., Amid C., Osanger A.,
RA  Fobos G., Han M., Wiemann S.;
RA  Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RL  SEQUENCE OF 11-959 FROM N.A. (ISOFORM 1).
RN  SEQUENCE OF 11-959 FROM N.A. (ISOFORM 1).
RC  TISSUE=Brain;
RX  MEDLINE=20039619; PubMed=10574462;
RA  Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA  Ohara O.;
RT  "Prediction of the coding sequences of unidentified human genes. XV.
RT  The complete sequences of 100 new cDNA clones from brain which code
RT  for large proteins in vitro."
RL  DNA Res. 6:337-345(1999).
RN  [3]
RP  PARTIAL SEQUENCE FROM N.A.
RX  MEDLINE=99038209; PubMed=9819391;
RA  Brown M.T., Andrade J., Radhakrishna H., Donaldson J.G., Cooper J.A.,
RA  Randazzo P.A.;
RT  "ASAP1, a phospholipid-dependent arf GTPase-activating protein that
RT  associates with and is phosphorylated by Src."
RL  Mol. Cell. Biol. 18:7038-7051(1998).
CC  -1- FUNCTION: Possesses phosphatidylinositol 4,5-bisphosphate-dependent
CC  GTPase-activating protein activity for ARF1 (ADP ribosylation
CC  factor 1) and ARF5 and a lesser activity towards ARF6. May
CC  coordinate membrane trafficking with cell growth or actin
CC  cytoskeleton remodeling by binding to both SRC and PIP2. May
CC  function as a signal transduction protein involved in the
CC  differentiation of fibroblasts into adipocytes and possibly other
CC  cell types (By similarity).
CC  -1- ENZYME REGULATION: Activity stimulated by phosphatidylinositol
CC  4,5-bisphosphate (PIP2) (By similarity).
CC  -1- SUBUNIT: Homodimer. Interacts with SRC and CRK (By similarity).
CC  -1- SUBCELLULAR LOCATION: Cytoplasmic; Predominantly. Membrane-
CC  associated; partially (By similarity).
CC  -1- ALTERNATIVE PRODUCTS:
CC  Event=Alternative splicing; Named isoforms=2;
CC  Name=2;
CC  IsoId=Q9ULH1-1; Sequence=Displayed;
CC  Name=1;

```



```

CC CC      IsoId=Q9ULH1-2; Sequence=VSP_008365;
CC CC      !- DOMAIN: The PH domain most probably contributes to the
CC CC      phosphoinositide-dependent regulation of ADP ribosylation factors
CC CC      (by similarity).
CC CC      !- PTM: Phosphorylated on tyrosine residues by SRC (By similarity).
CC CC      !- SIMILARITY: Contains 2 ANK repeats.
CC CC      !- SIMILARITY: Contains 1 Arf-GAP domain.
CC CC      !- SIMILARITY: Contains 1 C4-type zinc finger.
CC CC      !- SIMILARITY: Contains 1 PH domain.
CC CC      !- SIMILARITY: Contains 1 SH3 domain.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
DR DR      EMBL; BX537768; CAD97831.1; -;
DR DR      EMBL; AB033075; BAA86563.1; -;
DR DR      InAct; Q9ULH1; -;
DR DR      Genew; HGNC:2720; DDEF1.
DR DR      MIM; 605953; -;
DR DR      InterPro; IPR002110; ANK.
DR DR      InterPro; IPR001164; hrip like.
DR DR      InterPro; IPR000108; Neu_Cyt_fact_2.
DR DR      InterPro; IPR001849; PH.
DR DR      InterPro; IPR011036; PH_related.
DR DR      InterPro; IPR001452; SH3.
DR DR      Pfam; PF00023; ANK; 2.
DR DR      Pfam; PF01412; ArfGap; 1.
DR DR      Pfam; PF00169; PH; 1.
DR DR      Pfam; PF00018; SH3; 1.
DR DR      PRINTS; PR00499; P67PHOX.
DR DR      PRINTS; PR00405; REVINTACTNG.
DR DR      PRINTS; PR00452; SH3DOMAIN.
DR DR      ProDom; PD000066; SH3; 1.
DR DR      SMART; SM00248; ANK; 2.
DR DR      SMART; SM00105; ArfGap; 1.
DR DR      SMART; SM00233; PH; 1.
DR DR      SMART; SM00326; SH3; 1.
DR DR      PROSITE; PS00297; ANK_REPEAT; 1.
DR DR      PROSITE; PS00088; ANK_REPEAT; 1.
DR DR      PROSITE; PS01115; ARFGAP; 1.
DR DR      PROSITE; PS00003; PH_DOMAIN; 1.
DR DR      PROSITE; PS00002; SH3; 1.
DR DR      KW      Alternative splicing; ANK repeat; GTPase activation; Membrane; Repeat;
DR DR      SH3 domain; Zinc-finger.
FT FT      NON_TER 1 243 PH
FT FT      DOMAIN 151 243 Arf-GAP.
FT FT      ZN_FING 281 304 C4-type.
FT FT      REPEAT 427 459 ANK 1.
FT FT      REPEAT 463 492 ANK 2.
FT FT      DOMAIN 610 820 Pro-rich.
FT FT      DOMAIN 894 956 SH3.
FT FT      VARSP_LIC 130 130 E -> ESRR (in isoform 1).
FT FT      /FTID=VSP_008365.
SQ SEQUENCE 956 AA; 105967 MW; 6BCBB7034EC5BBE0 CRC64;
Query Match 2.2%; Score 22; DB 1; Length 956;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 444 FTWLSTNLGILTCIECSGHIHRE 465
Db 289 FTWLSTNLGILTCIECSGHIHRE 310
RESULT 13
Q8C1G6
ID Q8C1G6 PRELIMINARY; PRT; 123 AA.

```

```

AC Q8C1G6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
DE enriched library, clone:6030426L16 product:similar to KRUPPEL-RELATED
GN ZINC FINGER PROTEIN F80-L.
GN Name=6030426L16Rik;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaishwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multipillar sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK020054; BAC25612.1; -.
DR

```

```

DR MGD; MGI:1924364; 6030426116rik.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRA6.
DR Pfam; PF01352; KRA6; 1.
DR SMART; SM00349; KRA6; 1.
DR PROSITE; PS08085; KRA6; 1.
SQ SEQUENCE 123 AA; 14105 MW; 445173CF14A1A3AE CRC64;

Query Match 1.9%; Score 19; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 EDYKAPTASSFTTRTAQCR 34
Db 16 EDYKAPTASSFTTRTAQCR 34

RESULT 14
Q8R1L5 PRELIMINARY; PRT; 121 AA.
AC Q8R1L5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN CDNA 9030624G23 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young J.W., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska J., Smalios D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024416; AAH24416.1;
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRA6.
DR Pfam; PF01352; KRA6; 1.
DR SMART; SM00349; KRA6; 1.
DR PROSITE; PS08085; KRA6; 1.
SQ SEQUENCE 121 AA; 13904 MW; EC20C7A61B0634D1 CRC64;

Query Match 1.6%; Score 16; DB 2; Length 121;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MPDQISVSEFVAETHE 16
Db 1 MPDQISVSEFVAETHE 16

RESULT 15
Q9D308 PRELIMINARY; PRT; 155 AA.
AC Q9D308;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus adult male colon cDNA, RIKEN full-length enriched
DE library, clone:9030624G23 product:hypothetical protein, full insert
DE sequence.
GN Name=9030624G23Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=20493974; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata K., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.153600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multiplexed capillary sequencer."
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,

```

RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AK018572; BAB31284.1; -;
DR MGD; MGI:1914058; 9030624G23Rik.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR Pfam; PF01352; KRAB; 1
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS50805; KRAB; 1.
KW Hypothetical protein.
SQ SEQUENCE 155 AA; 17560 MW; D63034719828920E CRC64;

Query Match 1.5%; Score 15; DB 2; Length 155;
Best Local Similarity 100.0%; Pred.No. 3.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 APTASSFTTTRTAQCR 34
|||
Db 20 APTASSFTTTRTAQCR 34
|||

Search completed: August 5, 2005, 04:05:23
Job time : 121 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 02:52:18 ; Search time 31 Seconds
(without alignments)
3122.389 Million cell updates/sec

Title: US-09-914-042-1
Perfect score: 1006
Sequence: 1 MPDQISVSEFVAETHDYK.....DGDPRKGAPVSVFVFIAD 1006

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues
Word size : 0
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : PIR_79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	703	69.9	1006	2 T00050	hypothetical prote
2	30	3.0	1147	2 T42627	ADP-ribosylation f
3	12	1.2	1243	2 J55615	membrane-associate
4	10	1.0	138	2 G55096	hypothetical prote
5	9	0.9	903	2 T26743	hypothetical prote
6	9	0.9	923	2 T38398	hypothetical GTPas
7	9	0.9	951	2 T26738	hypothetical prote
8	9	0.9	975	2 T26737	hypothetical prote
9	9	0.9	1087	2 T49496	hypothetical prote
10	8	0.8	153	2 H75003	hypothetical prote
11	8	0.8	155	2 D82755	conserved hypothe
12	8	0.8	185	1 S60942	hypothetical prote
13	8	0.8	220	2 T26434	hypothetical prote
14	8	0.8	236	2 H87976	probable RING zinc
15	8	0.8	237	2 G84678	hypothetical prote
16	8	0.8	285	2 A75074	polyhedron envelop
17	8	0.8	312	1 JQ1559	polyhedral calyx p
18	8	0.8	313	2 T30486	hypothetical prote
19	8	0.8	320	2 G96714	transcription fact
20	8	0.8	359	2 S14283	probable pre-mRNA
21	8	0.8	398	2 T41600	UDP-N-acetylmuram
22	8	0.8	404	2 C69170	hypothetical prote
23	8	0.8	462	2 E70955	virB10 protein (im
24	8	0.8	482	2 E97748	cyclin a2-type, mi
25	8	0.8	484	2 T07675	hypothetical prote
26	8	0.8	539	2 D71260	hypothetical prote
27	8	0.8	596	2 T03908	hypothetical prote
28	8	0.8	627	2 T04562	conserved hypothe
29	8	0.8	678	2 T40362	

30	8	0.8	822	2 T01622	probable salt-indu
31	8	0.8	847	2 A82296	hypothetical prote
32	8	0.8	1292	2 D84727	probable RAD50 DNA
33	8	0.8	1948	2 B69511	N conserved hypoth
34	8	0.8	2100	2 T03223	probable polyketid
35	8	0.8	3712	2 S18253	laminin alpha-1 ch
36	8	0.8	4613	2 T17409	polyketide synthas
37	7	0.7	27	2 I57715	probasin - rat (fr
38	7	0.7	56	2 A98162	gene HE2 protein -
39	7	0.7	103	2 I37454	hypothetical prote
40	7	0.7	106	2 T00722	hypothetical prote
41	7	0.7	111	2 A53221	acidic ribosomal p
42	7	0.7	119	2 C40513	hypothetical prote
43	7	0.7	123	1 C44212	structural protein
44	7	0.7	123	2 G81127	hypothetical prote
45	7	0.7	127	2 A83861	aspartate 1-decarb

ALIGNMENTS

RESULT 1

T00050
hypothetical protein KIAA0400 - human
C:Species: Homo sapiens (man)
C>Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 05-Nov-1999
C;Accession: T00050
R;Ishikawa, K.; Nagase, T.; Nakajima, D.; Seki, N.; Ohira, M.; Miyajima, N.; Tanaka, A.,
submitted to the EMBL Data Library, October 1997
A;Description: Prediction of the coding sequences of unidentified human genes. VIII. The
A;Reference number: Z14080
A;Accession: T00050
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1006 <ISH>
A;Cross-references: EMBL:AB007860; NID:d1175316; PIDN:BA03696.1; PID:d1024577
A;Experimental source: brain, clone HG1091
C;Genetics:
A;Note: KIAA0400

Query Match 69.9%; Score 703; DB 2; Length 1006;				
Best Local Similarity 99.7%; Pred. No. 0;				
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
Qy	1	MPDQISVSEFVAETHDYKAPTASSFTRTAQCRNTVAAIEEALDVRMVLVYKMKSVKA	60	
Db	1	MPDQISVSEFVAETHDYKAPTASSFTRTAQCRNTVAAIEEALDVRMVLVYKMKSVKA	60	
Qy	61	INSSGLAHVENEQYTALEKFGGNCVCRDDPDGLSAPLKFSVFTKELTALFKNLIQNN	120	
Db	61	INSSGLAHVENEQYTALEKFGGNCVCRDDPDGLSAPLKFSVFTKELTALFKNLIQNN	120	
Qy	121	NIISFPLDSLKGDLKGVKDLKPPDKAWDYETKTKIEKEKEHAHKLHGMRTETLSG	180	
Db	121	NIISFPLDSLKGDLKGVKDLKPPDKAWDYETKTKIEKEKEHAHKLHGMRTETLSG	180	
Qy	181	ABIAEEMEKEKRRFQLOMCEVLLKVNELIKKGVDDLQNLIKYFHAQCNFFQDGLKAVES	240	
Db	181	ABIAEEMEKEKRRFQLOMCEVLLKVNELIKKGVDDLQNLIKYFHAQCNFFQDGLKAVES	240	
Qy	241	LKPSIETLSTDLHTIKQAQDEERQLIQLRDILKSALQVEQKESQIRQSTAYSILHQPG	300	
Db	241	LKPSIETLSTDLHTIKQAQDEERQLIQLRDILKSALQVEQKESQIRQSTAYSILHQPG	300	
Qy	301	NKEHGTNRNGSLYKKSQIRKQWQKRCVQNGFLTISHGTANRPPAKNLLTTCQVKNP	360	
Db	301	NKEHGTNRNGSLYKKSQIRKQWQKRCVQNGFLTISHGTANRPPAKNLLTTCQVKNP	360	
Qy	361	EEKKCFDLISHDRYTHFOADEQECQIWMVSLQNSKEEALNNAFGKDDNTGNNIVQSLT	420	
Db	361	EEKKCFDLISHDRYTHFOADEQECQIWMVSLQNSKEEALNNAFGKDDNTGNNIVQSLT	420	
Qy	421	KEIISVQRMVTGNDVCCDGPDPDTWLSTNLGILTCIECSGIHRELGVHYSPMQSLTLDV	480	

```
|||||
421 KEIISVQRMGTNDVCCDGPDPWLTSTNLGILTCIEGSHRELGVHYSRMQSLTLDV 480
Qy LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYTIERRYARKKH 540
Db LGTSELLAKNIGNAGFNEIMECCLPABDSVKPNPGSDMNAKDYITAKYTIERRYARKKH 540
Qy ADNAAKHLSCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHIAVSVDRTS 600
Db ADNAAKHLSCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETHLHIAVSVDRTS 600
Qy LHIIVDFVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGRKASIEIANESGETPLDI 660
Db LHIIVDFVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGRKASIEIANESGETPLDI 660
Qy AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLQSPENRRDRP 720
Db AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLQSPENRRDRP 720
Qy ISFYOLGSOIQSNAVSLARDAANLAKKORAFWPSILQNETYGALLSGSPPPQAPAPS 780
Db ISFYOLGSOIQSNAVSLARDAANLAKKORAFWPSILQNETYGALLSGSPPPQAPAPS 780
Qy TTSAPPLPPRVNKGKQVTASSANTLWKTNVSVDGGRQSSSDPPAVHPPLPLRLVSTN 840
Db TTSAPPLPPRVNKGKQVTASSANTLWKTNVSVDGGRQSSSDPPAVHPPLPLRLVSTN 840
Qy PLTTPPPPVAKTSVMEALSQSPKAPPGISQIRPPPLPPOPPSRLPQKPKAPGTDKST 900
Db PLTTPPPPVAKTSVMEALSQSPKAPPGISQIRPPPLPPOPPSRLPQKPKAPGADKST 900
Qy PLTNKGQPRGPDVLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCVADNP 960
Db PLTNKGQPRGPDVLSATEALGPLSNAMVLQPPAMPKRSQATKLKPKVKALYNCVADNP 960
Qy DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAFFVSFVHFAD 1006
Db DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAFFVSFVHFAD 1006

RESULT 2
T42627
ADP-ribosylation factor-directed GTPase activating protein, isoform a - mouse
N:Alternate names: ASAPla protein
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42627
R:Brown, M.T.; Andrade, J.; Radhakrishna, H.; Donaldson, J.G.; Cooper, J.A.; Randazzo, F.
Mol. Cell. Biol. 18, 7038-7051, 1998
A:Title: ASAP1, a phospholipid-dependent arf GTPase-activating protein that associates w
A:Reference number: 222178; MUID:99038209; PMID:9819391
A:Accession: T42627
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1147 <BRO>
A:Cross-references: UNIPROT:Q9QWY8; EMBL:AF075461; NID:g4063613; PID:g4063614; PIDN:AAC9
C:Genetics:
A:Gene: Shag1
C:Function:
A:Description: interacts directly with ADP-ribosylation factors (Arf1, Arfs and to a les
A:Note: supposed to coordinate membrane remodeling events
C:Keywords: membrane trafficking

Query Match 3.0%; Score 30; DB 2; Length 1147;
Best Local Similarity 100.0%; Pred. No. 1.6e-21;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LDSLLKGLKGVKGDLPKPKPKKAWDYETK 156
Db 147 LDSLLKGLKGVKGDLPKPKPKKAWDYETK 176

RESULT 3
```

```
JC5615
membrane-associated phosphatidyl inositol transfer protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 09-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 09-Jul-2004
C:Accession: JC5615
R:Aikawa, Y.; Hara, H.; Watanabe, T.
Biochem. Biophys. Res. Commun. 236, 559-564, 1997
A:Title: Molecular cloning and characterization of mammalian homologues of the Drosophil
A:Reference number: JC5615; MUID:97396134; PMID:9245688
A:Contents: Brain
A:Accession: JC5615
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1243 <AIK>
A:Cross-references: UNIPROT:O35954; DDBJ:AF006467
C:Comment: This protein plays a role in brain development, at a particular stage by tran
C:Genetics:
A:Gene: mpt-1
A:Map position: 19

Query Match 1.2%; Score 12; DB 2; Length 1243;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 772 PPAQPAAPSTTS 783
Db 811 PPAQPAAPSTTS 822

RESULT 4
G65096
hypothetical protein ygjM - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: G65096
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65096
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-138 <BLAT>
A:Cross-references: UNIPROT:P42594; GB:AE000390; GB:U00096; NID:g2367189; PIDN:AAC76117
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: YgJm
C:Superfamily: Escherichia coli hypothetical protein ygjM

Query Match 1.0%; Score 10; DB 2; Length 138;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 71 NEEQYTOALE 80
Db 26 NEEQYTOALE 35

RESULT 5
T26743
hypothetical protein Y39A1A.15c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26743
R:Wall, M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z20257
A:Accession: T26743
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-903 <WIL>
A:Cross-references: UNIPROT:Q9XX09; EMBL:AL031633; PIDN:CAA21032.1; GSPDB:GN00021; CBSP:
```

A;Experimental source: clone Y39A1A

C;Genetics:

A;Gene: CESP:Y39A1A.15c

A;Map position: 3

A;Introns: 33/3; 63/1; 139/3; 185/1; 382/3; 424/2; 798/3; 852/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 903;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 653 CIECSGIHR 661

RESULT 6

T38398

hypothetical GTPase activating protein for Arf protein - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C;Accession: T38398

R;McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, February 1996

A;Reference number: Z21791

A;Accession: T38398

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-923 <MCL>

A;Cross-references: UNIPROT:Q10165; EMBL:Z69240; PIDN:CAA93233.1; GSPDB:GN00066; SPDB:SF

A;Experimental source: strain 972h-; cosmid 26A3

C;Genetics:

A;Gene: SPDB:SPAC26A3.10

A;Map position: 1

A;Introns: 6/1; 47/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 923;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 751 CIECSGIHR 759

RESULT 7

T26738

hypothetical protein Y39A1A.15b - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T26738

R;Wall, M.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z20257

A;Accession: T26738

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-951 <WIL>

A;Cross-references: UNIPROT:Q9U2M2; EMBL:AL031633; PIDN:CAA21027.1; GSPDB:GN00021; CESP:

A;Experimental source: clone Y39A1A

C;Genetics:

A;Gene: CESP:Y39A1A.15b

A;Map position: 3

A;Introns: 34/3; 62/1; 81/3; 111/1; 187/3; 233/1; 430/3; 472/2; 846/3; 900/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 951;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 701 CIECSGIHR 709

RESULT 8

T26737

hypothetical protein Y39A1A.15a - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T26737

R;Wall, M.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z20257

A;Accession: T26737

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-975 <WIL>

A;Cross-references: UNIPROT:Q9XX14; EMBL:AL031633; PIDN:CAA21026.1; GSPDB:GN00021; CESP:

A;Experimental source: clone Y39A1A

C;Genetics:

A;Gene: CESP:Y39A1A.15a

A;Map position: 3

A;Introns: 105/3; 135/1; 211/3; 257/1; 454/3; 496/2; 870/3; 924/3

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 975;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 725 CIECSGIHR 733

RESULT 9

T49496

hypothetical protein B14D6.480 [imported] - Neurospora crassa

C;Species: Neurospora crassa

C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C;Accession: T49496

R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A;Reference number: Z25022

A;Accession: T49496

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1087 <SCH>

A;Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.480

A;Experimental source: BAC clone B14D6; strain OR74A

C;Genetics:

A;Gene: NCSP:B14D6.480

A;Map position: 6

A;Introns: 67/3; 99/3; 270/2; 589/2; 616/2

Query Match

Best Local Similarity 0.9%; Score 9; DB 2; Length 1087;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 CIECSGIHR 464

Db 811 CIECSGIHR 819

RESULT 10

H75003

hypothetical protein PAB1032 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C;Accession: H75003

R;Anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A;Reference number: A75001

A;Accession: H75003

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-153 <KAW>

A;Cross-references: UNIPROT:Q9UYE2; GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB5047

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1032

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB1032

Query Match 0.8%; Score 8; DB 2; Length 153;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 ITKIEKEK 164

|||||

Db 123 ITKIEKEK 130

RESULT 11

D82755

C:Species: conserved hypothetical protein XF0849 [imported] - Xylella fastidiosa (strain 9a5c)

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: D82755

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: D82755

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-155 <SIM>

A:Cross-references: UNIPROT:Q9PFR29; GB:AE003924; GB:AE003849; NID:g9105750; PIDN:AAF8365

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.B.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.Y.;

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0849

Query Match

Best Local Similarity 0.8%; Score 8; DB 2; Length 155;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 HGMIRTEI 178

|||||

Db 94 HGMIRTEI 101

RESULT 12

S60942

N:Alternate names: hypothetical protein YOR215c - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: S60942; S67108; S71717

R:Galissou, F.; Dujon, B.

submitted to the EMBL Data Library, October 1995

A:Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome

A:Reference number: S60938

A:Accession: S60942

A:Molecule type: DNA

A:Residues: 1-185 <GAL>

A:Cross-references: UNIPROT:Q12032; EMBL:X92441; NID:g1050762; PID:g1050767

R:Boyer, J.; Fairhead, C.; Gaillon, L.; Galissou, F.; Michaux, G.; Thierry, A.; Dujon, B

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67104

A:Accession: S67108

A:Molecule type: DNA

A:Residues: 1-185 <BOY>

A:Cross-references: EMBL:Z75123; NID:g1420500; PID:g1420501; GSPDB:GN00015; MIPS:YOR215c

A:Experimental source: strain S288C

R:Galissou, F.; Dujon, B.

Yeast 12, 877-885, 1996

A:Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV o

A:Reference number: S71713; MUID:96437977; PMID:18840505

A:Accession: S71717

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-185 <GAW>

A:Cross-references: EMBL:X92441; NID:g1050762; PIDN:CAA63178.1; PID:g1050767

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Gene: MIPS:YOR215c

A:Cross-references: SGD:S0005741

A:Map position: 1SR

C:Superfamily: Bacillus subtilis conserved hypothetical protein yqey

Query Match

Best Local Similarity 0.8%; Score 8; DB 1; Length 185;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 SLLKGD LK 136

|||||

Db 30 SLLKGD LK 37

RESULT 13

T26434

C:Species: hypothetical protein Y106G6H.14 - Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T26434

R:McMurray, A.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20214

A:Accession: T26434

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-220 <WIL>

A:Cross-references: UNIPROT:Q9U2Z5; EMBL:AL032631; PIDN:CAA21577.2; GSPDB:GN00019; CBSP:

A:Experimental source: clone Y106G6H

C:Genetics:

A:Gene: CESP:Y106G6H.14

A:Map position: 1

A:Introns: 32/3; 72/1; 143/3; 188/3

Query Match

Best Local Similarity 0.8%; Score 8; DB 2; Length 220;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 ELTFSEGD 969

|||||

Db 35 ELTFSEGD 42

RESULT 14

H87976

C:Species: protein Y106G6H.14 [imported] - Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C:Accession: H87976

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/c_elegans/ and www.sanger.ac.uk/Projects/C_ele

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: H87976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-236 <STO>
 A:Cross-references: UNIPROT:Q9U2Z5; GB:chr_I; PIDN:CAA21577.1; PID:G3880686; GSPDB:GN000
 C:Genetics:
 A:Gene: Y106G6H.14
 A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 236;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 ELTFSEGD 969
 |||||
 DB 35 ELTFSEGD 42

RESULT 15
 G84678
 Probable RING zinc finger protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
 C:Accession: G84678
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84678
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-237 <STO>
 A:Cross-references: UNIPROT:Q9SUJ7; GB:AE002093; NID:G4510422; PIDN:AAD21508.1; GSPDB:GN
 C:Genetics:
 A:Gene: At2G27940
 A:Map position: 2

Query Match 0.8%; Score 8; DB 2; Length 237;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 841 PLTPTPPP 848
 |||||
 DB 33 PLTPTPPP 40

Search completed: August 5, 2005, 04:05:56
 Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 04:05:29 ; Search time 104 Seconds
(without alignments)
3776.348 Million cell updates/sec

Title: US-09-914-042-1
Perfect score: 1006
Sequence: 1 MPDQISVSEFAETHEHYKA.....DGDGRKGAPVSVFVFIAD 1006

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 1752860 seqs, 390397842 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1752860

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Published Application#A:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10E_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	10.9	208	US-10-104-047-2284	Sequence 2284, App
2	110	10.9	350	US-09-764-875-635	Sequence 635, App
3	79	7.9	349	US-09-764-875-956	Sequence 956, App
4	30	3.0	136	US-10-276-774-1754	Sequence 1754, App
5	30	3.0	358	US-10-408-765A-2815	Sequence 2815, App
6	30	3.0	1132	US-10-210-281-74	Sequence 74, App
7	16	1.6	156	US-10-276-774-1638	Sequence 1638, App
8	13	1.3	43	US-09-864-761-46527	Sequence 46527, A
9	10	1.0	138	US-09-912-020-300	Sequence 300, App
10	10	1.0	138	US-10-771-241-300	Sequence 300, App
11	10	1.0	903	US-10-104-047-2951	Sequence 2951, App

Result No.	Score	Query Match	Length	ID	Description
12	10	1.0	903	US-10-490-605-2	Sequence 2, App
13	10	1.0	1358	US-10-778-804-11	Sequence 11, App
14	9	0.9	122	US-10-080-334-240	Sequence 240, App
15	9	0.9	125	US-10-437-963-156203	Sequence 156203, App
16	9	0.9	138	US-10-176-306-44	Sequence 44, App
17	9	0.9	326	US-10-369-493-2559	Sequence 2559, App
18	9	0.9	556	US-10-424-599-22658	Sequence 22658, App
19	9	0.9	580	US-10-094-749-2012	Sequence 2012, App
20	9	0.9	641	US-10-425-115-336077	Sequence 336077, App
21	9	0.9	686	US-10-221-625-6	Sequence 6, App
22	9	0.9	726	US-10-467-434-17	Sequence 17, App
23	9	0.9	759	US-10-080-334-236	Sequence 236, App
24	9	0.9	764	US-10-080-334-80	Sequence 80, App
25	9	0.9	778	US-10-080-334-238	Sequence 238, App
26	9	0.9	792	US-10-437-963-163744	Sequence 163744, App
27	9	0.9	804	US-10-080-334-235	Sequence 235, App
28	9	0.9	834	US-10-176-306-11	Sequence 11, App
29	9	0.9	834	US-10-080-334-237	Sequence 237, App
30	9	0.9	834	US-10-467-434-4	Sequence 4, App
31	9	0.9	836	US-10-719-993-842	Sequence 842, App
32	9	0.9	836	US-10-719-993-843	Sequence 843, App
33	9	0.9	836	US-10-719-993-844	Sequence 844, App
34	9	0.9	856	US-10-719-993-845	Sequence 845, App
35	9	0.9	870	US-10-322-281-358	Sequence 358, App
36	9	0.9	882	US-10-334-143-9	Sequence 9, App
37	8	0.8	17	US-09-880-149-61	Sequence 61, App
38	8	0.8	17	US-09-880-132-61	Sequence 61, App
39	8	0.8	17	US-10-345-281-61	Sequence 61, App
40	8	0.8	48	US-09-864-761-37629	Sequence 37629, A
41	8	0.8	59	US-10-424-599-180387	Sequence 180387, App
42	8	0.8	65	US-10-437-963-156713	Sequence 156713, App
43	8	0.8	71	US-09-349-954-21	Sequence 21, App
44	8	0.8	71	US-09-307-007-21	Sequence 21, App
45	8	0.8	71	US-10-673-708-21	Sequence 21, App

ALIGNMENTS

RESULT 1
US-10-104-047-2284
Sequence 2284, Application US/10104047
Publication No. US20030236392A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: NO. US20030236392A1 full length cDNA
FILE REFERENCE: H1-A0105
CURRENT APPLICATION NUMBER: US/10/104,047
CURRENT FILING DATE: 2002-03-25
PRIORITY APPLICATION NUMBER:
PRIORITY FILING DATE:
NUMBER OF SEQ ID NOS: 4096
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2284
LENGTH: 208
TYPE: PRT
ORGANISM: Homo sapiens
US-10-104-047-2284

Query Match 10.9%; Score 110; DB 15; Length 208;
Best Local Similarity 100.0%; Pred. No. 2.6e-96;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q#	D#	Sequence	Score
897	99	DKSTPLTNKQPGPVLDLSATLALGPIISNMVLOPPAPMRKSOATLK...KRVKALNYCV 956	10.06
99	99	DKSTPLTNKQPGPVLDLSATLALGPIISNMVLOPPAPMRKSOATLK...KRVKALNYCV 158	10.06
957	159	ADNPDELTFEGGVITVDGSEDESWTIGTIDGGRGAPVSVFVFIAD 200	10.06
159	159	ADNPDELTFEGGVITVDGSEDESWTIGTIDGGRGAPVSVFVFIAD 200	10.06

US-09-764-875-635
 : Sequence 635, Application US/09764875
 : Publication No. US20040018969A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosen et al.
 : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 : FILE REFERENCE: P1202
 : CURRENT APPLICATION NUMBER: US/09/764,875
 : CURRENT FILING DATE: 2001-01-17
 : Prior application data removed - consult PALM or file wrapper
 : NUMBER OF SEQ ID NOS: 1249
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO: 635
 : LENGTH: 350
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: SITE
 : LOCATION: (191)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-875-635

Query Match 10.9%; Score 110; DB 11; Length 350;
 Best Local Similarity 100.0%; Pred. No. 4,2e-96;
 Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 897 DKSTLTNKGQGRGVDSATLGLPLSNAMVLOPPAPMPKRSQATLKKPRKVALYVCV 956
 DB 241 DKSTLTNKGQGRGVDSATLGLPLSNAMVLOPPAPMPKRSQATLKKPRKVALYVCV 300
 QY 957 ADNPDELTFESGDVLIIVDGEDEQEWIIGHIDGDPGRKGAFFVSVFHFAD 1006
 DB 301 ADNPDELTFESGDVLIIVDGEDEQEWIIGHIDGDPGRKGAFFVSVFHFAD 350

RESULT 3
 US-09-764-875-956
 : Sequence 956, Application US/09764875
 : Publication No. US20040018969A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosen et al.
 : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 : FILE REFERENCE: P1202
 : CURRENT APPLICATION NUMBER: US/09/764,875
 : CURRENT FILING DATE: 2001-01-17
 : Prior application data removed - consult PALM or file wrapper
 : NUMBER OF SEQ ID NOS: 1249
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO: 956
 : LENGTH: 349
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: SITE
 : LOCATION: (28)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 : NAME/KEY: SITE
 : LOCATION: (65)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 : NAME/KEY: SITE
 : LOCATION: (66)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 : NAME/KEY: SITE
 : LOCATION: (178)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 : NAME/KEY: SITE
 : LOCATION: (190)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 : NAME/KEY: SITE
 : LOCATION: (270)
 : OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-764-875-956

Query Match 7.9%; Score 79; DB 11; Length 349;
 Best Local Similarity 100.0%; Pred. No. 2e-66;
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 928 VLOPPAPMPKRSQATLKKPRKVALYVCVADNPDELTFESGDVLIIVDGEDEQEWIIGHID 987
 DB 271 VLOPPAPMPKRSQATLKKPRKVALYVCVADNPDELTFESGDVLIIVDGEDEQEWIIGHID 330
 QY 988 GDEPRKGAFFVSVFHFAD 1006
 DB 331 GDEPRKGAFFVSVFHFAD 349

RESULT 4
 US-10-276-774-1754
 : Sequence 1754, Application US/10276774
 : Publication No. US20040053245A1
 : GENERAL INFORMATION:
 : APPLICANT: Hyseq, Inc.
 : APPLICANT: Tang, Y, Tom et al
 : TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
 : FILE REFERENCE: 21272-030
 : CURRENT APPLICATION NUMBER: US/10/276,774
 : CURRENT FILING DATE: 2002-11-18
 : PRIOR APPLICATION NUMBER: 09/560,875
 : PRIOR FILING DATE: 2000-04-27
 : PRIOR APPLICATION NUMBER: 09/496,914
 : PRIOR FILING DATE: 2000-02-03
 : NUMBER OF SEQ ID NOS: 2700
 : SOFTWARE: Custom
 : SEQ ID NO: 1754
 : LENGTH: 136
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1)..(136)
 : OTHER INFORMATION: Xaa = any amino acid or nothing
 US-10-276-774-1754

Query Match 3.0%; Score 30; DB 15; Length 136;
 Best Local Similarity 100.0%; Pred. No. 7.4e-20;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 LDSLLKGDLLKGVKGDLLKPPDKAMKDYETX 156
 DB 93 LDSLLKGDLLKGVKGDLLKPPDKAMKDYETX 122

RESULT 5
 US-10-408-765A-2815
 : Sequence 2815, Application US/10408765A
 : Publication No. US2004010474A1
 : GENERAL INFORMATION:
 : APPLICANT: Ghosh, Soumitra S.
 : APPLICANT: Fahy, John D.
 : APPLICANT: Zhang, Bing
 : APPLICANT: Gibson, Bradford W.
 : APPLICANT: Taylor, Steven W.
 : APPLICANT: Glenn, Gary M.
 : APPLICANT: Warrick, Dale E.
 : TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 : FILE REFERENCE: 660088.465
 : CURRENT APPLICATION NUMBER: US/10/408,765A
 : CURRENT FILING DATE: 2003-04-04
 : NUMBER OF SEQ ID NOS: 3077
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO: 2815
 : LENGTH: 358
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-10-408-765A-2815

Query Match 3.0%; Score 30; DB 16; Length 358;
Best Local Similarity 100.0%; Pred. No. 1.8e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LDSLLKGLKGVKDLKPKFKDKWKDYETK 156
Db 199 LDSLLKGLKGVKDLKPKFKDKWKDYETK 228

RESULT 6

US-10-210-281-74
; Sequence 74, Application US/10210281
; Publication No. US20040030096A1
; GENERAL INFORMATION:
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Zhong, Mei
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles E.
; APPLICANT: Ji, Weizhen
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Stone, David J.
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Casman, Stacie
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc L.

TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
TITLE OF INVENTION: THE SAME

FILE REFERENCE: 21402-416D
CURRENT APPLICATION NUMBER: US/10/210,281
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: 60/309,501
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 60/310,291
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: 60/361,775
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: 60/310,951
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/361,832
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: 60/311,292
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/311,979
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 60/312,203
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/313,201
PRIOR FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: 60/313,702
PRIOR FILING DATE: 2001-08-20
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 191
SOFTWARE: CuraSeqList version 0.1

SEQ ID NO 74

LENGTH: 1132

TYPE: PRT

ORGANISM: Homo sapiens

US-10-210-281-74

Query Match 3.0%; Score 30; DB 15; Length 1132;
Best Local Similarity 100.0%; Pred. No. 4.9e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 LDSLLKGLKGVKDLKPKFKDKWKDYETK 156

Db 147 LDSLLKGLKGVKDLKPKFKDKWKDYETK 176

RESULT 7

US-10-276-774-1638
; Sequence 1638, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1638
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-276-774-1638

Query Match 1.6%; Score 16; DB 15; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 450 NLGILTCIECSGIHRE 465

Db 3 NLGILTCIECSGIHRE 18

RESULT 8

US-09-864-761-46527
; Sequence 46527, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aesomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/235,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 46527
;; LENGTH: 43
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;;
;; OTHER INFORMATION: MAP TO AC009682.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
;; OTHER INFORMATION: EST HUMAN HIT: AL044307.1, EVALUE 2.00e-08
;; OTHER INFORMATION: SWISSPROT HIT: Q06846, EVALUE 6.10e+00
;; NAME/KEY: unsure
;; LOCATION: 14
US-09-864-761-46527

Query Match 1.3%; Score 13; DB 9; Length 43;
Best Local Similarity 100.0%; Pred. No. 0.0005; Mismatches 0; Indels 0; Gaps 0;
Matches 13; Conservative 0;

Qy 180 GAEIAEEMEKERR 192
|||
Db 23 GAEIAEEMEKERR 35

RESULT 9
US-09-912-020-300
;; Sequence 300, Application US/09912020
;; Patent No. US20020045592A1
;; GENERAL INFORMATION:
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Ohlsen, Kari L.
;; APPLICANT: Trawick, John
;; APPLICANT: Forsyth, R. Allyn
;; APPLICANT: Froelich, Jamie M.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard
;; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;; FILE REFERENCE: ELITRA 001DV1
;; CURRENT APPLICATION NUMBER: US/09/912,020
;; PRIOR FILING DATE: 2001-07-23
;; PRIOR APPLICATION NUMBER: 09/492,709
;; PRIOR FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: 60/117,405
;; PRIOR FILING DATE: 1999-01-27
;; NUMBER OF SEQ ID NOS: 485
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 300
;; LENGTH: 138
;; TYPE: PRT
;; ORGANISM: E. Coli
US-09-912-020-300

Query Match 1.0%; Score 10; DB 9; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.1; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Qy 71 NEEQYTOALE 80
|||
Db 26 NEEQYTOALE 35

RESULT 10
US-10-771-241-300
;; Sequence 300, Application US/10771241
;; Publication No. US20040241715A1
;; GENERAL INFORMATION:
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Forsyth, R. Allyn
;; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;; FILE REFERENCE: ELITRA 001C1
;; CURRENT APPLICATION NUMBER: US/10/771,241
;; CURRENT FILING DATE: 2004-02-03
;; PRIOR APPLICATION NUMBER: 09/492,709
;; PRIOR FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: 60/117,405
;; PRIOR FILING DATE: 1999-01-27
;; NUMBER OF SEQ ID NOS: 485
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 300
;; LENGTH: 138
;; TYPE: PRT
;; ORGANISM: E. Coli
US-10-771-241-300

Query Match 1.0%; Score 10; DB 16; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.1; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Qy 71 NEEQYTOALE 80
|||
Db 26 NEEQYTOALE 35

RESULT 11
US-10-104-047-2951
;; Sequence 2951, Application US/10104047
;; Publication No. US20030236392A1
;; GENERAL INFORMATION:
;; APPLICANT: HELIX RESEARCH INSTITUTE
;; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
;; FILE REFERENCE: H1-A0105
;; CURRENT APPLICATION NUMBER: US/10/104,047
;; CURRENT FILING DATE: 2002-03-25
;; PRIOR APPLICATION NUMBER:
;; PRIOR FILING DATE:
;; NUMBER OF SEQ ID NOS: 4096
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 2951
;; LENGTH: 903
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-104-047-2951

Query Match 1.0%; Score 10; DB 15; Length 903;
Best Local Similarity 100.0%; Pred. No. 5.7; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Qy 359 NPEEKKCFDL 368
|||
Db 356 NPEEKKCFDL 365

RESULT 12
US-10-490-605-2
;; Sequence 2, Application US/10490605
;; Publication No. US20050019768A1
;; GENERAL INFORMATION:

```
; APPLICANT: JAPAN AS REPRESENTED BY THE PRESIDENT OF THE UNIVERSITY OF TOKYO
; APPLICANT: ONCOTHERAPY SCIENCE, INC.
; TITLE OF INVENTION: HEPATOCELLULAR CARCINOMA-RELATED GENES AND POLYPEPTIDES, AND METH
; FILE OF INVENTION: FOR DETECTING HEPATOCELLULAR CARCINOMAS
; FILE REFERENCE: 25371-029NATI/SEN-A0121p-US
; CURRENT APPLICATION NUMBER: US/10/490,605
; CURRENT FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: USSN 60/324,261
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: CA
; PRIOR FILING DATE: 2002-08-23
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-490-605-2

Query Match          1.0%; Score 10; DB 17; Length 903;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 NPEKKKCFDL 368
    |||||
Db 356 NPEKKKCFDL 365

RESULT 13
US-10-778-804-11
; Sequence 11, Application US/10778804
; Publication No. US20040224388A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Turner, Geoffrey
; APPLICANT: Pollerman, Sarah E.
; APPLICANT: Memmott, Stephen D.
; TITLE OF INVENTION: Hyphal Growth in Fungi
; FILE REFERENCE: GC575-2
; CURRENT APPLICATION NUMBER: US/10/778,804
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: US 09/275,549
; PRIOR FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1358
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-778-804-11

Query Match          1.0%; Score 10; DB 16; Length 1358;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 839 TNPLTPTPPP 848
    |||||
Db 341 TNPLTPTPPP 350

RESULT 14
US-10-080-334-240
; Sequence 240, Application US/10080334
; Publication No. US20040002584A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shimkets, Richard A
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Malyankar, Uriel M.
```

```
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Casman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Patturajan, Meera
; APPLICANT: Gangolli, Esha A
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Liu, Xiaohong
; APPLICANT: Baumgartner, Jason C.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Spaderna, Steven K
; APPLICANT: Zerhusen, Bryan D
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-275
; CURRENT APPLICATION NUMBER: US/10/080,334
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,523
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/322,712
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/311,980
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/330,307
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/278,796
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: 60/281,521
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/276,677
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,595
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/270,220
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/274,295
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/318,526
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/286,548
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/291,765
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/270,797
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/276,400
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/270,810
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 240
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-080-334-240

Query Match          0.9%; Score 9; DB 15; Length 122;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 456 CIECSGIHR 464
    |||||
Db 36 CIECSGIHR 44

RESULT 15
US-10-437-963-156203
; Sequence 156203, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
```



```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 158203
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55894C.1.pep
US-10-437-963-156203
```

```
Query Match          0.9%; Score 9; DB 16; Length 125;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      876 PPPLPPQPP 884
        |||||
Db      17 PPPLPPQPP 25
```

```
Search completed: August 5, 2005, 04:16:00
Job time : 106 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 5, 2005, 03:45:33 ; Search time 237 Seconds
(without alignments)
4957.866 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 1006

Sequence: 1 MPDQISVSEFVAETHEDYK.....DGDPRKGAFPVSVHFIAID 1006

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 6959266 seqs, 1168006243 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6959266

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

- 1: /cgn2_6/ptodata/1/paa/US06_COMB.pcp.*
- 2: /cgn2_6/ptodata/1/paa/US06_COMB.pcp.*
- 3: /cgn2_6/ptodata/1/paa/US07_COMB.pcp.*
- 4: /cgn2_6/ptodata/1/paa/US080_COMB.pcp.*
- 5: /cgn2_6/ptodata/1/paa/US081_COMB.pcp.*
- 6: /cgn2_6/ptodata/1/paa/US082_COMB.pcp.*
- 7: /cgn2_6/ptodata/1/paa/US083_COMB.pcp.*
- 8: /cgn2_6/ptodata/1/paa/US084_COMB.pcp.*
- 9: /cgn2_6/ptodata/1/paa/US085_COMB.pcp.*
- 10: /cgn2_6/ptodata/1/paa/US086_COMB.pcp.*
- 11: /cgn2_6/ptodata/1/paa/US087_COMB.pcp.*
- 12: /cgn2_6/ptodata/1/paa/US088_COMB.pcp.*
- 13: /cgn2_6/ptodata/1/paa/US089_COMB.pcp.*
- 14: /cgn2_6/ptodata/1/paa/US090_COMB.pcp.*
- 15: /cgn2_6/ptodata/1/paa/US091_COMB.pcp.*
- 16: /cgn2_6/ptodata/1/paa/US092_COMB.pcp.*
- 17: /cgn2_6/ptodata/1/paa/US093_COMB.pcp.*
- 18: /cgn2_6/ptodata/1/paa/US094_COMB.pcp.*
- 19: /cgn2_6/ptodata/1/paa/US095_COMB.pcp.*
- 20: /cgn2_6/ptodata/1/paa/US096_COMB.pcp.*
- 21: /cgn2_6/ptodata/1/paa/US097A_COMB.pcp.*
- 22: /cgn2_6/ptodata/1/paa/US097B_COMB.pcp.*
- 23: /cgn2_6/ptodata/1/paa/US098_COMB.pcp.*
- 24: /cgn2_6/ptodata/1/paa/US099A_COMB.pcp.*
- 25: /cgn2_6/ptodata/1/paa/US099B_COMB.pcp.*
- 26: /cgn2_6/ptodata/1/paa/US100_COMB.pcp.*
- 27: /cgn2_6/ptodata/1/paa/US101_COMB.pcp.*
- 28: /cgn2_6/ptodata/1/paa/US102_COMB.pcp.*
- 29: /cgn2_6/ptodata/1/paa/US103_COMB.pcp.*
- 30: /cgn2_6/ptodata/1/paa/US104_COMB.pcp.*
- 31: /cgn2_6/ptodata/1/paa/US105_COMB.pcp.*
- 32: /cgn2_6/ptodata/1/paa/US106_COMB.pcp.*
- 33: /cgn2_6/ptodata/1/paa/US107_COMB.pcp.*
- 34: /cgn2_6/ptodata/1/paa/US108_COMB.pcp.*
- 35: /cgn2_6/ptodata/1/paa/US109_COMB.pcp.*
- 36: /cgn2_6/ptodata/1/paa/US110_COMB.pcp.*
- 37: /cgn2_6/ptodata/1/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1006	100.0	1006	24	US-09-914-042-1	Sequence 1, Appli
2	703	69.9	1006	22	US-09-791-537-142678	Sequence 142678,
3	703	69.9	1006	24	US-09-949-002-361	Sequence 361, App
4	703	69.9	1006	24	US-09-949-003C-2272	Sequence 2272, Ap
5	693	68.9	911	24	US-09-949-002-425	Sequence 425, App
6	693	68.9	911	24	US-09-949-003C-2770	Sequence 2770, Ap
7	693	68.9	990	27	US-10-170-205B-7820	Sequence 7820, Ap
8	693	68.9	990	27	US-10-170-205B-18115	Sequence 18115, A
9	270	26.8	909	37	US-60-452-680-18114	Sequence 18114, A
10	195	19.4	826	24	US-09-914-042-2	Sequence 2, Appli
11	151	15.0	278	22	US-09-791-537-31632	Sequence 31632, A
12	110	10.9	208	27	US-10-104-047-2284	Sequence 2284, Ap
13	110	10.9	350	1	PCT-US01-01332-635	Sequence 635, App
14	110	10.9	350	22	US-09-764-875-635	Sequence 635, App
15	79	7.9	349	1	PCT-US01-01332-956	Sequence 956, App
16	79	7.9	349	22	US-09-764-875-956	Sequence 956, App
17	43	4.3	138	24	US-09-914-042-7	Sequence 7, Appli
18	39	3.9	66	37	US-60-200-109-1231	Sequence 1231, Ap
19	31	3.1	40	37	US-60-181-996-1300	Sequence 1300, Ap
20	30	3.0	115	27	US-10-170-205B-3508	Sequence 3508, Ap
21	30	3.0	136	1	PCT-US01-03800A-1754	Sequence 1754, Ap
22	30	3.0	136	28	US-10-276-774-1754	Sequence 1754, Ap
23	30	3.0	195	28	US-10-221-279-11235	Sequence 11235, A
24	30	3.0	358	30	US-10-408-765-2815	Sequence 2815, Ap
25	30	3.0	358	30	US-10-408-765A-2815	Sequence 2815, Ap
26	30	3.0	358	37	US-60-389-987-2815	Sequence 2815, Ap
27	30	3.0	358	37	US-60-412-418-2815	Sequence 2815, Ap
28	30	3.0	1072	1	PCT-US02-36151-51	Sequence 51, Appli
29	30	3.0	1090	22	US-09-791-537-127954	Sequence 127954,
30	30	3.0	1122	33	US-10-796-280-999	Sequence 999, App
31	30	3.0	1122	35	US-10-995-561-705	Sequence 705, App
32	30	3.0	1122	37	US-60-568-219-358	Sequence 358, App
33	30	3.0	1129	1	PCT-US02-36151-44	Sequence 44, Appli
34	30	3.0	1129	22	US-09-791-537-26287	Sequence 26287, A
35	30	3.0	1129	33	US-10-796-280-1000	Sequence 1000, Ap
36	30	3.0	1129	35	US-10-995-561-706	Sequence 706, App
37	30	3.0	1129	37	US-60-568-219-359	Sequence 359, App
38	30	3.0	1132	28	US-10-210-281-74	Sequence 74, Appli
39	30	3.0	1147	22	US-09-791-537-127952	Sequence 127952,
40	29	2.9	47	37	US-60-169-840-6265	Sequence 6265, Ap
41	29	2.9	54	37	US-60-170-373-3689	Sequence 3689, Ap
42	27	2.7	27	37	US-60-160-209-3848	Sequence 3848, Ap
43	27	2.7	54	37	US-60-170-373-3889	Sequence 3889, Ap
44	22	2.2	123	28	US-10-221-279-9377	Sequence 9377, Ap
45	22	2.2	669	27	US-10-170-205B-22195	Sequence 22195, A

ALIGNMENTS

RESULT 1

US-09-914-042-1
; Sequence 1, Application US/09914042
; GENERAL INFORMATION:
; APPLICANT: SCHLESSINGER, JOSEPH
; APPLICANT: ANDREEV, JULIAN
; TITLE OF INVENTION: PYK2 BINDING PROTEINS
; FILE REFERENCE: 038602/1224
; CURRENT APPLICATION NUMBER: US/09/914.042
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: PCT/US00/04647
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/121,125
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1006
; TYPE: PRT

i ORGANISM: Homo sapiens
US-09-914-042-1

Query Match 100.0%; Score 1006; DB 24; Length 1006;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1006; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA 60
Qy 61 INSSGLAHVNEEOYTOALEKFGNCVCRDDPDGLGSFLKSFVTKELTALFKNLIQNMN 120
Db 61 INSSGLAHVNEEOYTOALEKFGNCVCRDDPDGLGSFLKSFVTKELTALFKNLIQNMN 120
Qy 121 NIISFPLDSLKGDGLKGVKGDLPKDFKAWKDYETKITKIEKKEHAKLHGMIRTEISG 180
Db 121 NIISFPLDSLKGDGLKGVKGDLPKDFKAWKDYETKITKIEKKEHAKLHGMIRTEISG 180
Qy 181 AEIAEMEKEKRRFFLOQCEYLLKVNELIKKGVLLQNLIKYFHAQCNFPDGLKAVES 240
Db 181 AEIAEMEKEKRRFFLOQCEYLLKVNELIKKGVLLQNLIKYFHAQCNFPDGLKAVES 240
Qy 241 LKPSIETLSTDLHTTIKQAQDEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSILHQPG 300
Db 241 LKPSIETLSTDLHTTIKQAQDEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSILHQPG 300
Qy - 301 NKEHGTERRNGSLYKKS DGIRKVKQKCSVKNGFLTISHGTANRPPAKNLILTCQVKTNP 360
Db 301 NKEHGTERRNGSLYKKS DGIRKVKQKCSVKNGFLTISHGTANRPPAKNLILTCQVKTNP 360
Qy 361 EKKCFDLISHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBLT 420
Db 361 EKKCFDLISHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBLT 420
Qy 421 KEIISEVQRMGTNDVCCCGAPDPTWLSNLGILTCIECSGIHRELGVHSPMQSLTLDV 480
Db 421 KEIISEVQRMGTNDVCCCGAPDPTWLSNLGILTCIECSGIHRELGVHSPMQSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNARKDYITAKYIERRYARKKH 540
Qy 541 ADNAKLHSLCEAVKTRDI FGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS 600
Db 541 ADNAKLHSLCEAVKTRDI FGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS 600
Qy 601 LHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Qy 661 AKRLKHEHCEBLLTQALSGRFSNVHVEYEWRLLEDLDESDDDMDEKLPQSENRRDRP 720
Db 661 AKRLKHEHCEBLLTQALSGRFSNVHVEYEWRLLEDLDESDDDMDEKLPQSENRRDRP 720
Qy 721 ISFYQLGNSQLQSNVSLARDANLAKQKRAFMPISILQNETYGALLSGSPPPAPAPAPS 780
Db 721 ISFYQLGNSQLQSNVSLARDANLAKQKRAFMPISILQNETYGALLSGSPPPAPAPAPS 780
Qy 781 TTSAPPLPRNVKQVQTASSANTLWKNTSVSVDGSGRQSSDDPVAHPPLPPLRVLTSTN 840
Db 781 TTSAPPLPRNVKQVQTASSANTLWKNTSVSVDGSGRQSSDDPVAHPPLPPLRVLTSTN 840
Qy 841 PLTPPPPPVAKTPSVMBSALQSPKAPPGISQIRPPLPPOPGRLLPQKPAAPTDKST 900
Db 841 PLTPPPPPVAKTPSVMBSALQSPKAPPGISQIRPPLPPOPGRLLPQKPAAPTDKST 900
Qy 901 PLTNKGQPRGPDVLSATEALGPLSNAMVLQPPAPMPKRSQATKLPKPKVLYNCVADNP 960
Db 901 PLTNKGQPRGPDVLSATEALGPLSNAMVLQPPAPMPKRSQATKLPKPKVLYNCVADNP 960
Qy 961 DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVPSFVHFIAID 1006
Db 961 DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVPSFVHFIAID 1006
```

Db 961 DELTFSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVPSFVHFIAID 1006

RESULT 2

US-09-791-537-142678
; Sequence 142678, Application US/09791537
; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 142678

; LENGTH: 1006

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-791-537-142678

Query Match 69.9%; Score 703; DB 22; Length 1006;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLVKMKSVKA 60
Qy 61 INSSGLAHVNEEOYTOALEKFGNCVCRDDPDGLGSFLKSFVTKELTALFKNLIQNMN 120
Db 61 INSSGLAHVNEEOYTOALEKFGNCVCRDDPDGLGSFLKSFVTKELTALFKNLIQNMN 120
Qy 121 NIISFPLDSLKGDGLKGVKGDLPKDFKAWKDYETKITKIEKKEHAKLHGMIRTEISG 180
Db 121 NIISFPLDSLKGDGLKGVKGDLPKDFKAWKDYETKITKIEKKEHAKLHGMIRTEISG 180
Qy 181 AEIAEMEKEKRRFFLOQCEYLLKVNELIKKGVLLQNLIKYFHAQCNFPDGLKAVES 240
Db 181 AEIAEMEKEKRRFFLOQCEYLLKVNELIKKGVLLQNLIKYFHAQCNFPDGLKAVES 240
Qy 241 LKPSIETLSTDLHTTIKQAQDEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSILHQPG 300
Db 241 LKPSIETLSTDLHTTIKQAQDEERRQLIQLRDILKSALQVEKEDSQIRQSTAYSILHQPG 300
Qy 301 NKEHGTERRNGSLYKKS DGIRKVKQKCSVKNGFLTISHGTANRPPAKNLILTCQVKTNP 360
Db 301 NKEHGTERRNGSLYKKS DGIRKVKQKCSVKNGFLTISHGTANRPPAKNLILTCQVKTNP 360
Qy 361 EKKCFDLISHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBLT 420
Db 361 EKKCFDLISHDRTHYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBLT 420
Qy 421 KEIISEVQRMGTNDVCCCGAPDPTWLSNLGILTCIECSGIHRELGVHSPMQSLTLDV 480
Db 421 KEIISEVQRMGTNDVCCCGAPDPTWLSNLGILTCIECSGIHRELGVHSPMQSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPADS VKPNPGSDMNARKDYITAKYIERRYARKKH 540
Qy 541 ADNAKLHSLCEAVKTRDI FGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS 600
Db 541 ADNAKLHSLCEAVKTRDI FGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRTS 600
Qy 601 LHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHIVDFLVQNSGNDLKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Qy 661 AKRLKHEHCEBLLTQALSGRFSNVHVEYEWRLLEDLDESDDDMDEKLPQSENRRDRP 720
Db 661 AKRLKHEHCEBLLTQALSGRFSNVHVEYEWRLLEDLDESDDDMDEKLPQSENRRDRP 720
```

```
Db 661 AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRREDP 720
Qy 721 ISFYQLGNSQLQSNVSLARDAANLAKEQKRAFMPSTLQNETYCALLSGSPPPAQAAPS 780
Db 721 ISFYQLGNSQLQSNVSLARDAANLAKEQKRAFMPSTLQNETYCALLSGSPPPAQAAPS 780
Qy 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSRORSSSDPPAVHPPPLPLRVSTN 840
Db 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSRORSSSDPPAVHPPPLPLRVSTN 840
Qy 841 PLTPTPPPPVAKTSPVMEALSQPSKAPPGISQIRPPPLPQPPSRLLPQKPACTDKST 900
Db 841 PLTPTPPPPVAKTSPVMEALSQPSKAPPGISQIRPPPLPQPPSRLLPQKPACTDKST 900
Qy 901 PLTNKGQPRGVDLSATEALGFLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVNP 960
Db 901 PLTNKGQPRGVDLSATEALGFLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVNP 960
Qy 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFHFAD 1006
Db 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFHFAD 1006

RESULT 3
US-09-949-002-361
; Sequence 361, Application US/09949002
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 361
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Human
; US-09-949-002-361

Query Match 69.9%; Score 703; DB 24; Length 1006;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
Qy 61 INSSGLAHVENEQYTOALEFGGNCVCRDDPDGLSAPLKFVSFTKLTALFKNLIONMN 120
Db 61 INSSGLAHVENEQYTOALEFGGNCVCRDDPDGLSAPLKFVSFTKLTALFKNLIONMN 120
Qy 121 NIISFPLDSLKGLDKGVGDLKPFDKAWKDYETKITKEKEKEHAKLHGMRTISG 180
Db 121 NIISFPLDSLKGLDKGVGDLKPFDKAWKDYETKITKEKEKEHAKLHGMRTISG 180
Qy 181 AEIAEEMEKERRFFLOQCEYLLKWEIKIKKGVDDLQNLKIFYHAQCNFFQDGLKAVES 240
Db 181 AEIAEEMEKERRFFLOQCEYLLKWEIKIKKGVDDLQNLKIFYHAQCNFFQDGLKAVES 240
Qy 241 LKPSIETLSTDLHTIKQAQDERQLIOLRILKLSALQVEQKDSQIRQSTAYSILHQPOG 300
Db 241 LKPSIETLSTDLHTIKQAQDERQLIOLRILKLSALQVEQKDSQIRQSTAYSILHQPOG 300
Qy 301 NKEHGTENRGLYKKSQDGIKRWKQKCSVKNQGLFTITSHGTANRPPAKNLLTQCVKTNP 360
Db 301 NKEHGTENRGLYKKSQDGIKRWKQKCSVKNQGLFTITSHGTANRPPAKNLLTQCVKTNP 360
Qy 361 BEKCFDLISHDRTYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBLT 420
```

```
Db 361 BEKCFDLISHDRTYHFOAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGNNIVQBLT 420
Qy 421 KEIISVQRMGTNDVCCDCCGAPDPTLSTNLGILTCIECSGIHRELGVHVSQMSLTLDV 480
Db 421 KEIISVQRMGTNDVCCDCCGAPDPTLSTNLGILTCIECSGIHRELGVHVSQMSLTLDV 480
Qy 481 LGTSELALLAKNIGNAGFNEIMECCLPABDSVKPNPFGSDMMARKDYITAKYIERYARKKH 540
Db 481 LGTSELALLAKNIGNAGFNEIMECCLPABDSVKPNPFGSDMMARKDYITAKYIERYARKKH 540
Qy 541 ADNAKHLHSLCEAVKTRDIFGLLOAYADGVLDLTKIPLANGHEDEETALHLAVRSVDRTS 600
Db 541 ADNAKHLHSLCEAVKTRDIFGLLOAYADGVLDLTKIPLANGHEDEETALHLAVRSVDRTS 600
Qy 601 LHIIVDFLVQNSGNDLKOTGKSTALHYCCLTDNAECLLRLRGKASIEIANESGETPLDI 660
Db 601 LHIIVDFLVQNSGNDLKOTGKSTALHYCCLTDNAECLLRLRGKASIEIANESGETPLDI 660
Qy 661 AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRREDP 720
Db 661 AKRLKHECBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRREDP 720
Qy 721 ISFYQLGNSQLQSNVSLARDAANLAKEQKRAFMPSTLQNETYCALLSGSPPPAQAAPS 780
Db 721 ISFYQLGNSQLQSNVSLARDAANLAKEQKRAFMPSTLQNETYCALLSGSPPPAQAAPS 780
Qy 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSRORSSSDPPAVHPPPLPLRVSTN 840
Db 781 TTSAPPLPRNVGKQVQTASSANTLWKTNSVSDGSRORSSSDPPAVHPPPLPLRVSTN 840
Qy 841 PLTPTPPPPVAKTSPVMEALSQPSKAPPGISQIRPPPLPQPPSRLLPQKPACTDKST 900
Db 841 PLTPTPPPPVAKTSPVMEALSQPSKAPPGISQIRPPPLPQPPSRLLPQKPACTDKST 900
Qy 901 PLTNKGQPRGVDLSATEALGFLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVNP 960
Db 901 PLTNKGQPRGVDLSATEALGFLSNAMVLQPPAMPKRSQATKLKPKRVKALYNADVNP 960
Qy 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFHFAD 1006
Db 961 DELTFSEGDVLIIVDGEEDQEWIIGHIDGDPGRKGAFFVSVFHFAD 1006

RESULT 4
US-09-949-003C-2272
; Sequence 2272, Application US/09949003C
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPSYCHIATRIC
; TITLE OF INVENTION: DISORDERS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000791
; CURRENT APPLICATION NUMBER: US/09/949,003C
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,446
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 74065
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2272
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-949-003C-2272

Query Match 69.9%; Score 703; DB 24; Length 1006;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1003; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKS VKA 60
Qy 61 INSSGLAHVENEQYTOALEFGGNCVCRDDPDGLSAPLKFVSFTKLTALFKNLIONMN 120
```

```
|||||
Db 61 INSSGLAHVNEEQYTQALEKFGNCVCRDDPDGLGSAPLKESVFTKELTALFKNLQNMN 120
Qy 121 NIISFPDLSLLKGLDKGVKGDLPKFPDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 180
Db 121 NIISFPDLSLLKGLDKGVKGDLPKFPDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 180
Qy 181 AEIAEEMEKEKRRFFQLOQCEVLLKNEIKIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 240
Db 181 AEIAEEMEKEKRRFFQLOQCEVLLKNEIKIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 240
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 300
Db 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 300
Qy 301 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNVGLTISHGTANRPPAKNLNLLTCQVKTNP 360
Db 301 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNVGLTISHGTANRPPAKNLNLLTCQVKTNP 360
Qy 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGENNIVQELT 420
Db 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGENNIVQELT 420
Qy 421 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 480
Db 421 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNARKDYITAKYIERRYARKKH 540
Qy 541 ADNAAKHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600
Db 541 ADNAAKHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600
Qy 601 LHIYDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHIYDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Qy 661 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 720
Db 661 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 720
Qy 721 ISFYQLGNSQLQSNNAVLARDAANLAKKQAFMPSILQNETYGALLSGSPPPQAPAPS 780
Db 721 ISFYQLGNSQLQSNNAVLARDAANLAKKQAFMPSILQNETYGALLSGSPPPQAPAPS 780
Qy 781 TTSAAPLPPRVNKGKVTQASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 840
Db 781 TTSAAPLPPRVNKGKVTQASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 840
Qy 841 PLTPPTPPPVAKTPSVMBALSQSPKAPPGISQIRPPPLPPOPPRLPQKPAAGTKST 900
Db 841 PLTPPTPPPVAKTPSVMBALSQSPKAPPGISQIRPPPLPPOPPRLPQKPAAGTKST 900
Qy 901 PLTNKGQPRGVDLSATEALGPLSNAMVLOPPAMPKRSQATKLKPKVKALYNCAVNP 960
Db 901 PLTNKGQPRGVDLSATEALGPLSNAMVLOPPAMPKRSQATKLKPKVKALYNCAVNP 960
Qy 961 DELTSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVSVFVFIAD 1006
Db 961 DELTSEGDVIIVDGEEDQEWIIGHIDGDPGRKGAPVSVFVFIAD 1006
```

RESULT 5

```
US-09-949-002-425
; Sequence 425, Application US/09949002
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
```

```
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 425
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Human
US-09-949-002-425

Query Match 68.9%; Score 693; DB 24; Length 911;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MPDQISVSSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLVKMKSVKA 60
Db 17 MPDQISVSSEFVAETHEDYKAPTASSFTTRTAQCRNTVAAIEEALDVRMVLVKMKSVKA 76
Qy 61 INSSGLAHVNEEQYTQALEKFGNCVCRDDPDGLGSAPLKESVFTKELTALFKNLQNMN 120
Db 77 INSSGLAHVNEEQYTQALEKFGNCVCRDDPDGLGSAPLKESVFTKELTALFKNLQNMN 136
Qy 121 NIISFPDLSLLKGLDKGVKGDLPKFPDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 180
Db 137 NIISFPDLSLLKGLDKGVKGDLPKFPDKAWKDYETKITKEKEKEHAKLHGMIRTEISG 196
Qy 181 AEIAEEMEKEKRRFFQLOQCEVLLKNEIKIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 240
Db 197 AEIAEEMEKEKRRFFQLOQCEVLLKNEIKIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 256
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 300
Db 257 LKPSIETLSTDLHTIKQAQDEERRQLIQLRDLILKSALQVEQKESQIRQSTAYSILHOPOG 316
Qy 301 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNVGLTISHGTANRPPAKNLNLLTCQVKTNP 360
Db 317 NKEHGTERNGSLYKKSQDGIKRWQKRCVKNVGLTISHGTANRPPAKNLNLLTCQVKTNP 376
Qy 361 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGENNIVQELT 420
Db 377 EEKCFDLISHDRTYHFQAEDEQECQIWMVSLQNSKEEALNNAFKGDDNTGENNIVQELT 436
Qy 421 KEIISVQRMGTGNDVCCDGPDPFWLSTNLGILTCIECSGIHRELGVHSPMOSLTLVDV 480
Db 437 KEIISVQRMGTGNDVCCDGPDPFWLSTNLGILTCIECSGIHRELGVHSPMOSLTLVDV 496
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 497 LGTSELLAKNIGNAGFNEIMECCLPADSVPKPNPGSDMNARKDYITAKYIERRYARKKH 556
Qy 541 ADNAAKHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 600
Db 557 ADNAAKHSLCEAVKTRDI FGLLOAYADGVDLTEKIPLANGHEPDETALHLAVSRVDRTS 616
Qy 601 LHIYDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 617 LHIYDFLVQNSGNDKQTKGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 676
Qy 661 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 720
Db 677 AKRLKHECEBELLTQALSGRNSHVHVEYEWRLHEDLDESDDDMDEKLPSPNRRDRP 736
Qy 721 ISFYQLGNSQLQSNNAVLARDAANLAKKQAFMPSILQNETYGALLSGSPPPQAPAPS 780
Db 737 ISFYQLGNSQLQSNNAVLARDAANLAKKQAFMPSILQNETYGALLSGSPPPQAPAPS 796
Qy 781 TTSAAPLPPRVNKGKVTQASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 840
Db 797 TTSAAPLPPRVNKGKVTQASSANTLWKTNSVSDGSRQSRSSDPPPAVHPPLPLRVSTN 856
Qy 841 PLTPPTPPPVAKTPSVMBALSQSPKAPPGISQIRPPPLPPOPPRLPQKPAAGTKST 895
```

Db 857 PLTPTPPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 911
|||||
RESULT 6
US-09-949-003C-2770
; Sequence 2770, Application US/09949003C
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPSYCHIATRIC
; TITLE OF INVENTION: DISORDERS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000791
; CURRENT APPLICATION NUMBER: US/09/949,003C
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,446
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 74065
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2770
; LENGTH: 911
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-949-003C-2770
Query Match 68.9%; Score 693; DB 24; Length 911;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKSKVKA 60
Db 17 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKSKVKA 76
Qy 61 INSSGLAHVNEEYOYTALEKFGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLIQNMN 120
Db 77 INSSGLAHVNEEYOYTALEKFGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLIQNMN 136
Qy 121 NIISFPLDSLKGLKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 180
Db 137 NIISFPLDSLKGLKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 196
Qy 181 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 240
Db 197 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 256
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYSILHQPQ 300
Db 257 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYSILHQPQ 316
Qy 301 NKEHGTNRNGSLYKSDGIRKWKRCVKNQGLTIKSHGTANRPPAKNLITCOVKTNP 360
Db 317 NKEHGTNRNGSLYKSDGIRKWKRCVKNQGLTIKSHGTANRPPAKNLITCOVKTNP 376
Qy 361 EEKCFDLISHDRTYHFQADEQECQIWMVLSKEEALNNAFKGDDNTGNNIVBELT 420
Db 377 EEKCFDLISHDRTYHFQADEQECQIWMVLSKEEALNNAFKGDDNTGNNIVBELT 436
Qy 421 KEIISVQRMGTNDVCCDCCGAPDPTWLSNLGILTCIECSGIHRELGVHVSQMSLTLDV 480
Db 437 KEIISVQRMGTNDVCCDCCGAPDPTWLSNLGILTCIECSGIHRELGVHVSQMSLTLDV 496
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 497 LGTSELLAKNIGNAGFNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYARKKH 556
Qy 541 ADNAKHLCEAVKTRIDFGLQYADGVDLTEKIPLANGHEPDETALHLAVSVDRTS 600
Db 557 ADNAKHLCEAVKTRIDFGLQYADGVDLTEKIPLANGHEPDETALHLAVSVDRTS 616
Qy 601 LHIIVDFVQNSGLDKQTKGSTALHYCCLTDNAECLLKGKASIBIANESGETPLDI 660
Db 617 LHIIVDFVQNSGLDKQTKGSTALHYCCLTDNAECLLKGKASIBIANESGETPLDI 676
Qy 661 AKRLKHECEBELLTQALSGRFNSHVHVEYEWRLLEDLDESDDDMDEKQLQPSNNRDRP 720

Db 677 AKRLKHECEBELLTQALSGRFNSHVHVEYEWRLLEDLDESDDDMDEKQLQPSNNRDRP 736
Qy 721 ISFYQLGSNQLQSNVSLARDAANLAKEKQRAFMPSTILQNETYCALISGSPPPAQPAAPS 780
Db 737 ISFYQLGSNQLQSNVSLARDAANLAKEKQRAFMPSTILQNETYCALISGSPPPAQPAAPS 796
Qy 781 TTSAPPLPPRVNGKVQTASSANTLWKTNSVSDGSGRQRSSSDPPAVHPPLPPLRVVTSTN 840
Db 797 TTSAPPLPPRVNGKVQTASSANTLWKTNSVSDGSGRQRSSSDPPAVHPPLPPLRVVTSTN 856
Qy 841 PLTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 895
Db 857 PLTTPPPVAKTSPVMEALSQSPKAPPGISQIRPPPLPQPPSRLPQKKPAPG 911
RESULT 7
US-10-170-205E-7820
; Sequence 7820, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7820
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-7820
Query Match 68.9%; Score 693; DB 27; Length 990;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKSKVKA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKSKVKA 60
Qy 61 INSSGLAHVNEEYOYTALEKFGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLIQNMN 120
Db 61 INSSGLAHVNEEYOYTALEKFGNCVCRDDPDGLGSAPLKFVSFTKELTALFKNLIQNMN 120
Qy 121 NIISFPLDSLKGLKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 180
Db 121 NIISFPLDSLKGLKGVKGLKPPDKAMKDYETKITKEKEKEHAKLHGMIRTEISG 180
Qy 181 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 240
Db 181 AEIAEEMEKERRFPQLOMCEYLLKVNELIKKGVDDLQNLIKYFHAQCNFPQDGLKAVES 240
Qy 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYSILHQPQ 300
Db 241 LKPSIETLSTDLHTIKQAQDEERRQLIQRLDILKSALQVEKEDSQIRQSTAYSILHQPQ 300
Qy 301 NKEHGTNRNGSLYKSDGIRKWKRCVKNQGLTIKSHGTANRPPAKNLITCOVKTNP 360
Db 301 NKEHGTNRNGSLYKSDGIRKWKRCVKNQGLTIKSHGTANRPPAKNLITCOVKTNP 360
Qy 361 EEKCFDLISHDRTYHFQADEQECQIWMVLSKEEALNNAFKGDDNTGNNIVBELT 420
Db 361 EEKCFDLISHDRTYHFQADEQECQIWMVLSKEEALNNAFKGDDNTGNNIVBELT 420
Qy 421 KEIISVQRMGTNDVCCDCCGAPDPTWLSNLGILTCIECSGIHRELGVHVSQMSLTLDV 480
Db 421 KEIISVQRMGTNDVCCDCCGAPDPTWLSNLGILTCIECSGIHRELGVHVSQMSLTLDV 480
Qy 481 LGTSELLAKNIGNAGFNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYARKKH 540
Db 481 LGTSELLAKNIGNAGFNEIMECCLPADSVKPNPGSDMNARKDYITAKYIERRYARKKH 540

```
QY 541 ADNAKLSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRS 600
Db 541 ADNAKLSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRS 600
QY 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
QY 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLLEDHEDDESDDMDKQSPNRRDRP 720
Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLLEDHEDDESDDMDKQSPNRRDRP 720
QY 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
Db 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
QY 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
Db 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
QY 841 PLTPTPPPPVAKTPSVMEALSQPSKPAPPGISQIRPPPLPQPPSRLPQKKPAPG 895
Db 841 PLTPTPPPPVAKTPSVMEALSQPSKPAPPGISQIRPPPLPQPPSRLPQKKPAPG 895

RESULT 8
US-60-452-680-18115
; Sequence 18115, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18115
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-18115

Query Match 68.9%; Score 693; DB 37; Length 990;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 893; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKSVA 60
Db 1 MPDQISVSEFVAETHEDYKAPTASSFTTRTAQCRNTVAABEALDVRMVLKMKKSVA 60
QY 61 INSSGLAHVENEEOYTOALEKFGNCVCRDPPDLGSAFLKFSVFTKELTALFKNLQNMN 120
Db 61 INSSGLAHVENEEOYTOALEKFGNCVCRDPPDLGSAFLKFSVFTKELTALFKNLQNMN 120
QY 121 NIISPLSLKGLDKGVKGLDKKPFDKAWKDYETKITIKIEKKEKHAHLGMIIRTEISG 180
Db 121 NIISPLSLKGLDKGVKGLDKKPFDKAWKDYETKITIKIEKKEKHAHLGMIIRTEISG 180
QY 181 AEIAEEMEKERRFFOLQWCEVLLKYNELIKKGVLDLQNLIKYFHAQCNFQDGLKAVES 240
Db 181 AEIAEEMEKERRFFOLQWCEVLLKYNELIKKGVLDLQNLIKYFHAQCNFQDGLKAVES 240
QY 241 LKPSITETLSTDLHTTKQADEERRQLIQRLDILKSALQVEQKESQIRQSTAYSILHQPG 300
Db 241 LKPSITETLSTDLHTTKQADEERRQLIQRLDILKSALQVEQKESQIRQSTAYSILHQPG 300
QY 301 NKEHGTENSGLYKKSQDGIKRWQKRCVKNQFLTISHGTANRPPAKNLLTTCQVKNP 360
Db 301 NKEHGTENSGLYKKSQDGIKRWQKRCVKNQFLTISHGTANRPPAKNLLTTCQVKNP 360
```

```
QY 361 BEKKCFDLISHDRTHYFQAEDEQECQIWMSVLQNSKEEALNNAFKGDDTGNENIVBELT 420
Db 361 BEKKCFDLISHDRTHYFQAEDEQECQIWMSVLQNSKEEALNNAFKGDDTGNENIVBELT 420
QY 421 KEIISVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSIHRELGVHYSMPQSLTLDV 480
Db 421 KEIISVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSIHRELGVHYSMPQSLTLDV 480
QY 481 LGTSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMMARKDYITAKYIERYARKKH 540
Db 481 LGTSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMMARKDYITAKYIERYARKKH 540
QY 541 ADNAKLSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRS 600
Db 541 ADNAKLSLCEAVKTRDIFGLQAYADGVDLTEKIPLANGHEPDETALHLAVRSVDRS 600
QY 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
Db 601 LHI VDFLVQSGNLDKQTGKSTALHYCCLTDNAECLKLLRGKASIEIANESGETPLDI 660
QY 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLLEDHEDDESDDMDKQSPNRRDRP 720
Db 661 AKRLKHEHCEBELLTQALSGRNSHVHVEYEWRLLEDHEDDESDDMDKQSPNRRDRP 720
QY 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
Db 721 ISFYOLGSGNOLQOSNAVSLARDAANLAKKORAFMPSILQNETYGNALLSGSPPPAQAAPS 780
QY 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
Db 781 TTSAPPLPRNVGKVQTASSANTLWKTNSVSDGSGRQSSSDPPAVHPPPLPLRVSTN 840
QY 841 PLTPTPPPPVAKTPSVMEALSQPSKPAPPGISQIRPPPLPQPPSRLPQKKPAPG 895
Db 841 PLTPTPPPPVAKTPSVMEALSQPSKPAPPGISQIRPPPLPQPPSRLPQKKPAPG 895

RESULT 9
US-60-452-680-18114
; Sequence 18114, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18114
; LENGTH: 909
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-18114

Query Match 26.8%; Score 270; DB 37; Length 909;
Best Local Similarity 99.7%; Pred. No. 4e-257;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 342 ANRPPAKNLLTTCQVKTNPBEKCFDLISHDRTHYFQAEDEQECQIWMSVLQNSKEEALN 401
Db 306 ANRPPAKNLLTTCQVKTNPBEKCFDLISHDRTHYFQAEDEQECQIWMSVLQNSKEEALN 365
QY 402 NAFKGGDDNTGNNIVQBELTKEIISVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSG 461
Db 366 NAFKGGDDNTGNNIVQBELTKEIISVQRTMGNDVCCGAPDPTWLSTNLGILTCIECSG 425
QY 462 IHRELGVHYSMPQSLTLDVLTGTSSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMNA 521
Db 426 IHRELGVHYSMPQSLTLDVLTGTSSELLAKNIGNAGNEIMECCLPADS VKPNPGSDMNA 485
QY 522 RKDYITAKYIERYARKKHADNAKLSLCEAVKTRDIFGLQAYADGVDLTEKIPLANG 581
```


Db 486 RKDYITAKYIERRARKHADAALSLCSAVKTRDIFGLQAYADGVDTLTKIPLANG 545
QY 582 HEPDETALHLAVRSVDRSLHIVDFLVQNSGNLDKQTKGSGTALHYCCCLTDNAECLKLL 641
Db 546 HEPDETALHLAVRSVDRSLHIVDFLVQNSGNLDKQTKGSGTALHYCCCLTDNAECLKLL 605
QY 642 RGKASIEIANSGETPLDIAKRLKHEHCEELLTQALSGRFSNHHVVEYEWRLLEDDES 701
Db 606 RGKASIEIANSGETPLDIAKRLKHEHCEELLTQALSGRFSNHHVVEYEWRLLEDDES 665
QY 702 DDDWDEKLQPS 712
Db 666 DDDWDEKLQPS 676

RESULT 10

US-09-914-042-2

; Sequence 2, Application US/09914042
; GENERAL INFORMATION:
; APPLICANT: SCHLESINGER, JOSEPH
; APPLICANT: ANDREEV, JULIAN
; TITLE OF INVENTION: PVK2 BINDING PROTEINS
; FILE REFERENCE: 038602/1224
; CURRENT APPLICATION NUMBER: US/09/914,042
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: PCT/US00/04647
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/121,125
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-914-042-2

Query Match 19.4%; Score 195; DB 24; Length 826;

Best Local Similarity 100.0%; Pred. No. 7.7e-183;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 LYKSDGTRKWKQKCSVKGNGFLTISHGTANRPAPKUNLITCQVKNPBEKKCFDLISH 371
Db 140 LYKSDGTRKWKQKCSVKGNGFLTISHGTANRPAPKUNLITCQVKNPBEKKCFDLISH 199
QY 372 DRTYHFOADEQECQIWMVQLNSKEEALNNAFKGDDNTGNNIVQELTKIISVQVMT 431
Db 200 DRTYHFOADEQECQIWMVQLNSKEEALNNAFKGDDNTGNNIVQELTKIISVQVMT 259
QY 432 GNDVCCDCGADPTWLSTNLGILTCIECSGIHRELGVHYSQMOSLTLDLVLTSELLAKN 491
Db 260 GNDVCCDCGADPTWLSTNLGILTCIECSGIHRELGVHYSQMOSLTLDLVLTSELLAKN 319
QY 492 IGNAGFNEIMBCCLP 506
Db 320 IGNAGFNEIMBCCLP 334

RESULT 11

US-09-791-537-31632

; Sequence 31632, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31632
; LENGTH: 278
; TYPE: PRT
; ORGANISM: pdb 1DCQA
US-09-791-537-31632

Query Match 15.0%; Score 151; DB 22; Length 278;
Best Local Similarity 100.0%; Pred. No. 1.1e-139;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 544 AAKLHSLCEAVKTRDIFGLQAYADGVDTLTKIPLANGHEPDETALHLAVRSVDRSLH 603
Db 128 AAKLHSLCEAVKTRDIFGLQAYADGVDTLTKIPLANGHEPDETALHLAVRSVDRSLH 187
QY 604 VDFLVQNSGNLDKQTKGSGTALHYCCCLTDNAECLKLLRGKASIEIANSGETPLDIAKR 663
Db 188 VDFLVQNSGNLDKQTKGSGTALHYCCCLTDNAECLKLLRGKASIEIANSGETPLDIAKR 247
QY 664 LKHEHCEELLTQALSGRFSNHHVVEYEWRL 694
Db 248 LKHEHCEELLTQALSGRFSNHHVVEYEWRL 278

RESULT 12

US-10-104-047-2284

; Sequence 2284, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2284
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2284

Query Match 10.9%; Score 110; DB 27; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-99;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 897 DKSTPLTNKGGPRGVDLSATEALGPLSNAMVLOPPAMPBKSOATKLKPKRVKALYN 956
Db 99 DKSTPLTNKGGPRGVDLSATEALGPLSNAMVLOPPAMPBKSOATKLKPKRVKALYN 158
QY 957 ADNPELTFSEGDIIVDGEDQEWIGHIDGDPGRKGAFVSVFHFIA 1006
Db 159 ADNPELTFSEGDIIVDGEDQEWIGHIDGDPGRKGAFVSVFHFIA 208

RESULT 13

PCT-US01-01332-635

; Sequence 635, Application PC/TUS0101332
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PUZ02PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01332
; CURRENT FILING DATE: 2001-05-09
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 635
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE

; LOCATION: (191)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01332-635

Query Match 10.9%; Score 110; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No. 5.7e-99;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 897 DKSTPLTNKGQPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKLPKRVKALYNVCV 956
Db 241 DKSTPLTNKGQPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKLPKRVKALYNVCV 300
Qy 957 ADNPDPLTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFVPSFVHFIAAD 1006
Db 301 ADNPDPLTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFVPSFVHFIAAD 350

RESULT 14
US-09-764-875-635
; Sequence 635 Application US/09764875
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 635
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (191)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-635

Query Match 10.9%; Score 110; DB 22; Length 350;
Best Local Similarity 100.0%; Pred. No. 5.7e-99;
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 897 DKSTPLTNKGQPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKLPKRVKALYNVCV 956
Db 241 DKSTPLTNKGQPRGPDLSATEALGPLSNAMVLQPPAPMPRKSOATKLPKRVKALYNVCV 300
Qy 957 ADNPDPLTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFVPSFVHFIAAD 1006
Db 301 ADNPDPLTFSEGDVIVDGEEDQEWIIGHIDGDPGRKGAFVPSFVHFIAAD 350

RESULT 15
PCT-US01-01332-956
; Sequence 956 Application PC/TUS0101332
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01332
; CURRENT FILING DATE: 2001-05-09
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 956
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE

; LOCATION: (65)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (66)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (178)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (190)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (270)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01332-956

Query Match 7.9%; Score 79; DB 1; Length 349;
Best Local Similarity 100.0%; Pred. No. 3e-68;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 928 VLQPPAPMPRKSOATKLPKRVKALYNVCVADNPDELTFSEGDVIVDGEEDQEWIIGHID 987
Db 271 VLQPPAPMPRKSOATKLPKRVKALYNVCVADNPDELTFSEGDVIVDGEEDQEWIIGHID 330
Qy 988 GDCGRKGAFVPSFVHFIAAD 1006
Db 331 GDCGRKGAFVPSFVHFIAAD 349

Search completed: August 5, 2005, 04:10:01
Job time : 241 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 08:32:11 ; Search time 7497 Seconds
(without alignments)
5107.730 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYK.....DGDGPKGAPVPSFVHFAD 1006

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgm2_1/USPTO_spool_h/US0914042/runat_04082005_090632_2122/app_query.fasta_1.1159
-DB=EST -QFMT=fastcap -SUFFIX=first -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=spt -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0914042 @CGN 1 1 4326 @runat_04082005_090632_2122 -NCPU=6 -ICPU=3
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

```
EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsel.*
9: gb_gsel.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1592.5	30.2	1167	5	BUS08193
2	1409.5	26.7	884	5	BQ439046
3	1391	26.4	824	7	CNS35522
4	1352.5	25.6	909	5	BUI159073
5	1351	25.6	773	7	CN460643
6	1351	25.6	781	7	CO429713
7	1271	24.1	731	7	CNS38273
8	1271	24.1	742	7	CNS525450
9	1269	24.1	901	5	BUI172807

10	1266	24.0	784	6	CA328666	UI-M-FY0-
11	1260	23.9	742	7	CNS33008	UI-M-H00-
12	1246.5	23.6	740	5	BM944240	UI-M-EH0P
13	1237	23.5	738	7	CNS37253	UI-M-HS0-
14	1235	23.4	737	7	CF539222	UI-M-GH0-
15	1229	23.3	861	5	BUI22224	603488266
16	1228.5	23.3	846	5	BUI55022	AGENCOURT
17	1226	23.3	761	7	CF535252	UI-M-GH0-
18	1215	23.0	689	5	BUI703510	UI-M-F00-
19	1211	23.0	977	5	BQ882484	AGENCOURT
20	1196.5	22.7	753	7	CF535194	UI-M-GH0-
21	1196	22.7	742	7	CNS35077	UI-M-HS0-
22	1188	22.5	702	7	CF532249	UI-M-GH0-
23	1177.5	22.3	1014	4	BG293568	602390218
24	1174	22.3	891	5	BUI10830	AGENCOURT
25	1171	22.2	735	6	CB246523	UI-M-F10-
26	1155	21.9	714	7	CF745192	UI-M-GV0-
27	1143.5	21.7	801	5	EX914137	EX914137
28	1140	21.6	680	6	BY720163	BY720163
29	1132.5	21.5	682	4	B1562457	603256345
30	1129.5	21.4	780	5	EX914135	EX914135
31	1103	20.9	663	7	CR769039	DKF20468J
32	1092.5	20.7	684	2	BE617384	601442018
33	1091	20.7	913	5	BQ422593	AGENCOURT
34	1089.5	20.7	948	6	CD358314	AGENCOURT
35	1083.5	20.5	3115	3	BC038615	Mus muscu
36	1062	20.1	629	7	CN165820	996655 MA
37	1041	19.7	702	6	CD350751	UI-M-G10-
38	1039.5	19.7	924	5	BUI253672	603746385
39	1031	19.6	612	7	CF539480	UI-M-GH0-
40	1031	19.6	741	6	CD494693	CD412-C03
41	1008.5	19.1	775	5	BUI32648	603120421
42	995.5	18.9	978	4	BM471456	AGENCOURT
43	982	18.6	795	7	CR435039	CR435039
44	977	18.5	567	7	CF735384	UI-M-HB0-
45	968	18.4	965	5	BUI151927	AGENCOURT

ALIGNMENTS

RESULT 1
BUS08193
LOCUS BUS08193 1167 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10128476 NIH_MGC_71 Homo sapiens CDNA clone IMAGE:6502231
5', mRNA sequence.
ACCESSION BUS08193
VERSION BUS08193.1 GI:22814426
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1167)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL4057 row: p column: 08
High quality sequence stop: 762.
Location/Qualifiers
1. 1167
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6502231"

/tissue type="leiomyosarcoma"
 /lab host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.1 kb."

ORIGIN

Alignment Scores:
 Pred. No.: 2,98e-104 Length: 1167
 Score: 1592.50 Matches: 328
 Percent Similarity: 77.73% Conservative: 14
 Best Local Similarity: 74.55% Mismatches: 38
 Query Match: 30.20% Indels: 60
 DB: 5 Gaps: 6

US-09-914-042-1 (1-1006) x BU508193 (1-1167)

Qy 404 PheLysGlyAspAspAsnThr-GlyGluAsnAsnIleValGlnGlnLeuThrLysGluI 423
 Db 2 TTTAAGGGGATGACAACTACTGGGAGAAATACATGCTCCAGAACTGACAAAGGAGAT 61
 Qy 423 eileSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAs 443
 Db 62 CATCTCAGAAGTGCAGAGATGACGGGCAATGACGCTCTGTGACTGTGGGGGCCAGA 121
 Qy 443 pProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHi 463
 Db 122 TCCTACATGGCTTTCCACCACTGGGCATCTGACCTGCATCGAGTGTTCGGGAATCCA 181
 Qy 463 sArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyTh 483
 Db 182 CCGAGAGCTGGGGTTCATATTATCCAGATGCAGTCCCTGACCTTAGATGACTGGGAAC 241
 Qy 483 rSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCy 503
 Db 242 ATCTGAGCTGCTGCTCGCAAGATATTGGGAATGCAGGCTTTAATGAGATCATGGAATG 301
 Qy 503 sCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgL 523
 Db 302 TTGCTTACAGCTGAGGACTCAGTCCAAACCCCAACCCAGCGACATGAATGCAAGAAA 361
 Qy 523 sAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHisAlaAspAs 543
 Db 362 CGACTACATCAGCCCAAGTATCATCGAGGAGATACCGAAGGAAGACGACCGGATPA 421
 Qy 543 nAlaLysLeuHisSerLeuGlyAlaValLysThrArgAspIlePheGlyLeuLe 563
 Db 422 CGCGCGAAGCTTACAGTCTTTGCGAGCGCTCAAAACGAGAGATATTTTGGATTGCT 481
 Qy 563 uGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnGlyHisG 583
 Db 482 CCAAGCTTATGCTGATGTGTGTGATCTTACGGAAAAAATCCCACTGGCCAAACGGACATGA 541
 Qy 583 uProAspGluThrAlaLeuHisLeuAlaValArgSerValAspAspThrSerLeuHisI 603
 Db 542 GCCGGATGAACGGCCCTCCACCTTGCAGTTCAGATCCGTTGATCGAATCTCTTCACAT 601
 Qy 603 eValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerTh 623
 Db 602 TGTAGACTTTTATGTTTCAGACAGTGGGAACCTGGATAAACAGACAGAGGAAAGGCAGCAC 661
 Qy 623 rAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgG 643
 Db 662 AGCCCTGCATCTACTGCTGCTGACCGCAATGTCGAGTGCCTCAAGTTCCTCTCGGGG 721
 Qy 643 YLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysAr 663
 Db 722 GAAGGCTTCCATCGAGATAGCAACAGTCCAGAGAGACTCCGCTGGACATTCGCCAAGCG 781
 Qy 663 gLeuLysHisGluHisCysGluGluLeuThrGlnAlaLeuSerGlyArgPheAsnSe 683
 Db 782 CCTCAAGCACGAGCACTGTGAGGAGCTGTGACCCCGAGCTTATCTGGAAAAATTTAAATTC 841

Qy 683 rHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGluSerAspAs 703
 Db 842 TCAGTTCCTCCCTTGAATATATATGCGACTACTCCAGAAAAACCGGATGAAGTATGA 901
 Qy 703 pAspMetAspGlu-LysLeuGlnProSerGlu-AsnArgArg---GluAspArgProIle 721
 Db 902 CGACTCGATGAAAAAATTGACGCCAGTTCCTCAACCGGGGGAACACGGCCCTTC 961
 Qy 722 SerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAsp 741
 Db 962 GCCTTCTACCGCGTGGG- 979
 Qy 742 AlaAlaAsnLeuAlaLysGluLysGluArgAlaPheMetProSerIleLeuGlnAsnGlu 761
 Db 979 ----- 979
 Qy 762 ThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAla----- 777
 Db 980 -----CCTCCAACCAACTTTCAGCCTAAACCTGTAA 1012
 Qy 778 -----AlaProSerThrThrSerAlaProLeuProProArgAsnValGlyLysVal 795
 Db 1013 TCTTTGGGCCCAAAA---ATTGCTGGCCAAACCTTGCCCAAGAAAAACCAAAAGGG 1069
 Qy 796 GlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGly 815
 Db 1070 GCTTTTTCCTGCCCCAGCAATCTTTGGCCAAAATTTGAAACCTTAAC-----GGG 1120
 Qy 816 SerArgGlnArgSerSerAspProProAlaValHisProProLeuProPro 833
 Db 1121 AGCCCCCTCTGGAGAGGGCAGAGCCCCC-----CCTCCCGGCCCCCA 1165

RESULT 2
 BQ439046
 LOCUS
 DEFINITION AGNCOURT 7761599 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6020058
 5' mRNA sequence.
 ACCESSION BQ439046
 VERSION BQ439046.1 GI:21178122
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 884)
 NIH-MGC http://mgs.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Straubeberg, Ph.D.
 Email: csapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM13223 row: i column: 19
 High quality sequence stop: 733.
 FEATURES
 Location/Qualifiers
 1..884
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6020058"
 /tissue_type="epithelioid carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_70"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."

ORIGIN

Alignment Scores:

Pred. No.: 2,996-91 Length: 884
Score: 1409.50 Matches: 279
Percent Similarity: 95.55% Conservative: 0
Best Local Similarity: 95.55% Mismatches: 7
Query Match: 26.73% Indels: 6
DB: 5 Gaps: 1

US-09-914-042-1 (1-1006) x BQ439046 (1-884)

Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 13 GGGATCCGAAAGTGTGGCAGAAAAGGAATGTTTCAGTTAAATATGGTTTCTGACCAT 72
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 73 TCCCATGGTACCGTAAACCGGCTCTGCAAGCTCAACCTGCTTAACCTGCAGGTGAAG 132
Qy 358 ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe 377
Db 133 ACCAACCTCGAGGAGAAAGTCTTTCACCTTATTTCCATGACAGAACTTACCACTTT 192
Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
Db 193 CAAGCTGAAGATGAACAGGAATGTCAAAATATGGATGTCTGTCTGCAAAATAGCAAGAA 252
Qy 398 GluAlaLeuAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417
Db 253 GAAGCTTTAAACAATGCATTTTAAGGGGGATGACAACTACTGGAGAAAATAAATCGTCCAA 312
Qy 418 GluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
Db 313 GAACGTCAAAAGGAGATCATCTCAGAAGTCAGAGGATGACGGCAATGACGTCTGCTGT 372
Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457
Db 373 GACTGTGGGGGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATC 432
Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477
Db 433 GAGTGTCCGGAATCCACGAGAGTGGGGGTTCATTATTCAGGATGACGTCCCTGACC 492
Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497
Db 493 TTAGATGTACTGGGNACATCTGAGCTGTCTCGCCAGAAATATGGGAATGCAGGCTTT 552
Qy 498 AsnGluIleMetGluCysLeuProAlaGluAspSerValLysProAsnProGlySer 517
Db 553 AATGAGATCATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGGCAGC 612
Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537
Db 613 GACATGAATCAAGAAAGAGCTACATCAGCCAAAGTACATCGAGAGAGATACCCAAAG 672
Qy 538 LysLysHisAlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557
Db 673 AAGAAGCAGCGGATACCGCGGAGCTTACAGTCTTTGCGAGCCCGCTCAAAACGAGA 732
Qy 558 AspIlePhe-GlyLeuLeuGlnAlaTyrAlaAspGlyVal-AspLeuThrGluLysIleP 577
Db 733 GATATTTTNGGATTTGCTTCAAGCTTATGCTGATGGTGGGATCTTACGGAAAAAATCC 792
Qy 577 roLeuAlaAsnGlyHisGluPro-AspGluThrAla--LeuHisLeuAlaValArgServ 596
Db 793 CACTGGCCAAACGACATGAGCGGGATGAACGGCCCTCCACACCTTGCAGTCAGAAATCC 852
Qy 596 alaAsp---ArgThrSerLeuHisIle 603
Db 853 GTGGATTGGAANCCCTCTCTTACATT 878

RESULT 3

CN535522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cga@bcm.tmc.edu

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mousefl.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. 824

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30670487"

/tissue_type="Upper Head"

/dev_stage="embryo 9.5 - 10.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH BMAP_HSO"

/note="Organ: Upper Head; Vector: pYX-Asc; Site: 1; EcoR I; Site 2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with Not I and then cloned

directionally into pYX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is CGAACTGAAT. This library was created for the University

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:

Pred. No.: 5,81e-90 Length: 824

Score: 1391.00 Matches: 269

Percent Similarity: 98.55% Conservative: 2

Best Local Similarity: 97.82% Mismatches: 4

Query Match: 26.38% Indels: 1

DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN535522 (1-824)

Qy 214 ValAspLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAsp 233

Db 1 GTGGATTGCTTCAGAAATCTGATCAAGTACTTTTCCGCCCATGTCATTTTTCAGGAT 60

Qy 234 GlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHis 253

Db 61 GGATTGAAGCAGTAGAAAGCTTCAAGCTTCCATCGAGACGCTCTCCACGACCTCCAC 120

Qy 254 ThrIleLysGlnAlaGlnAspGluGluArgGlnGlnLeuIleGlnLeuArgAspIleLeu 273

```

Db      121  ACCATCAACAGGCCAGGATGAGGAACGGCAGCAGCTGATACAACTTCGAGATATTTTG 180
Qy      274  LysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgGlnSerThrAlaTyr 293
Db      181  AAGTCAGCATTTGCAGGTGGAAAGAGAAAGAGACTCACAACTTCGCCAAAGTACAGCCTAC 240
Qy      294  SerLeuHisGlnProGlnGlnLysGluHisGlyThrGluArgAsnGlySerLeuTyr 313
Db      241  AGCTTACATCAACTCAGGGAACAGAGGAGCATGGACAGAGAGGAACCGAAGCTCTAC 300
Qy      314  LysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGly 333
Db      301  AAGAAGACGATGGATCCGAAAGTGTGGCAGAGAGAGAGTTCGTTAAATAATGGC 360
Qy      334  PheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuThr 353
Db      361  TTCCTCACCATCTCCACGGCACTGCCAACCGGCCACCTGCCAAGCTCAACCTGCTAAC 420
Qy      354  CysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArg 373
Db      421  TGCACAGTGAAGACCAATCTCTGAGGAGAAAGAGTGTTCGACCTCATATCATGACAGG 480
Qy      374  ThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGln 393
Db      481  ACGTACCACTTCCAAAGCGAAGACGAACAAGATGTCCAGATATGATGCTGTACTGCAG 540
Qy      394  AsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsn 413
Db      541  AACAGCAAGSAGAAAGCTCTGAACAAACGCCCTTTAAGGGTGTATGACAACACTGGAGAAAT 600
Qy      414  AsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsn 433
Db      601  AACATGTCCTCAAGAGCTGACCAAGAGATCATCTCGAGGTGCAGAGATGACGGGCAT 660
Qy      434  AspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIle 453
Db      661  GACGTGTCTGCGACTGTGGGCACA-GATTTCGACGTGGCTCTCTACCAACTGNGCATC 719
Qy      454  LeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMet 473
Db      720  CTGACTTGATCGAGTGTCTGTGGGATCCACCGGAGCTGGGGGTTCATTACTCCAGGATG 779
Qy      474  GlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeu 488
Db      780  CAGTCCCTGACATAGATGATTTGGGAACGCTGTGAGCTTCTGCTT 824

```

RESULT 4

```

BUI59073
LOCUS      BUI59073
DEFINITION AGENCOURT 7938808 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6012546
5', mRNA sequence.
ACCESSION BUI59073
VERSION    BUI59073.1 GI:22672983
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13203 row: p column: 19
High quality sequence stop: 636.

```

FEATURES

```

source
1..909
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6012546"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

```

ORIGIN

```

Alignment Scores:
Pred. No.:      3,95e-87      Length:      909
Score:          1352.50      Matches:    277
Percent Similarity: 93.07%      Conservative: 5
Best Local Similarity: 91.42%      Mismatches: 14
Query Match:     25.65%      Indels:     7
DB:              5          Gaps:        2
US-09-914-042-1 (1-1006) x BUI59073 (1-909)

Qy      385  CysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeuAsnAsnAlaPhe 404
Db      1   TGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAGAAGAGCTTTAAACAATGCATTT 60
Qy      405  LysGlyAspAspAsnThr-GlyGluAsnAsnIleValGlnGluLeuThrLysGluIleI 424
Db      61  AAGGGGATGACATATCTGGGAGAAATATACATCGTCCAGAACTGCACAAAGAGATCAT 120
Qy      424  eSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspPr 444
Db      121  CTCAGAAAGTCAGAGGATGACGGCAATGACGCTGTGCTGCTGCTGGGGCCGAGATCC 180
Qy      444  oThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisAr 464
Db      181  TACATGGCTTTCCACCAACCTGGGCATCCCTGACCTGATCGAGTTCGCGAATCCACCG 240
Qy      464  gGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSe 484
Db      241  AGAGCTGGGGTTCATTATTCCAGGATGCAGTCCCTGACCTTAGATGCTACTGGGAACATC 300
Qy      484  rGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCy 504
Db      301  TGAGCTGCTGCTCGCCCAAGAATATTTGGGAATGCAAGGCTTTAATGAGATCATGGAATGTTG 360
Qy      504  sLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAs 524
Db      361  CCTACCAGCTGAGACTCAGTCAAAACCCAAACCCAGGCGAGCATGATTAATGCAAAAGGA 420
Qy      524  pTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAl 544
Db      421  CTACATCAGCCCAAGTACATCGAGAGAGATACGCAAGGAGAAAGACCGCGGATACGC 480
Qy      544  aAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGl 564
Db      481  GGCGAAGCTTCACAGTCTTTCCGAGGCGCTCAAAACAGAGAGATATTTTGGATGCTCCA 540
Qy      564  nAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluPr 584
Db      541  AGCTTATGCTGATGTTGATCTTTCGGAATAAATCCCACTGGCCCAACGACATGAGCC 600
Qy      584  oAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSerLeuHisIleVa 604
Db      601  GGATGAACGGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCCTCTCTTCACATTGG 660
Qy      604  lAspPheLeuValGlnAsnSer-GlyAsnLeuAspLysGlnThrGlyLysGlySerThrA 624
Db      661  AGACTTTTGTAGTTCAAGAACAGTGGGAACTCTGGATAAACAGACAGAGGAAAGGCAGCAG 720

```

QY 624 laLeuHisTyrCysCysLeuThrAspAenAla-GluCysLeuLeuLeuLeuAsgly 643
 Db 721 CCCTGCACTACTGCTGCTGACCGACATGCCGAGGCCTCAAGTCTCTCTGCGGGG 780
 QY 644 Lys-AlaSerIleGluIleAlaAenGluSerGly---GluThrProLeuAsePilleAlaLy 662
 Db 781 AAGGCGCTCCATCGATACCAAGAGTCCGGAGAGACTCCGCTGGGACATGCCCA 840
 QY 662 sArgLeuHisGluHisCysGluGlu-----LeuLeuThrGlnAlaLeuSerGlyAr 680
 Db 841 GCGCCCTCAAGCACCAGCACTGGTGAAGAGCTTCTTTCACCAAGGCTTTATCTGGAAG 900
 QY 680 g 680
 Db 901 A 901

RESULT 5
 CN460643
 LOCUS
 DEFINITION UI-M-HB0-coy-a-05-0-UI.r1 NIH_BMAP_HB0 Mus musculus cDNA clone
 IMAGE:30652708 5', mRNA sequence.

ACCESSION CN460643
 VERSION CN460643.1 GI:46466369

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

JOURNAL

COMMENT

1 (bases 1 to 773)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: ccapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. 773
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30652708"
 /tissue_type="whole eye"
 /dev_stage="embryo 12.5, 13.5, 14.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_HB0"
 /note="Organ: Eye; Vector: pYX-Asc; Site 1: Ecor I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with Ecor I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TTATGAGT. This library was created for the University
 Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
 Developing Mouse Nervous System, supported by National
 Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:

Pred. No.:

4.02e-87

Length:

773

Score: 1351.00 Matches: 250
 Percent Similarity: 98.44% Conservativeness: 3
 Best Local Similarity: 97.28% Mismatches: 4
 Query Match: 25.62% Indels: 0
 DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN460643 (1-773)

QY 295 LeuHisGlnProGlnGlyAenLysGluHisGlyThrGluAraAenGlySerLeuTyrLys 314
 Db 2 TTACATCAACTCAGGGAACCAAGGAGCATGGAACAGAGAGAAACCGGAACCTCTTACAAG 61
 QY 315 LysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAenGlyPhe 334
 Db 62 AAGAGCGATGGGATCCGGAAGAGTGGCAGAGAGAGAGTGTTCGGTTAAATGGCTTC 121
 QY 335 LeuThrIleSerHisGlyThrAlaAenArgProAlaLysLeuAenLysLeuLeuTyrCys 354
 Db 122 CTCACCATCTCCACAGGCACTGCCAACCGGCACCTGCCAAGCTCAACTGCTAACCTGC 181
 QY 355 GlnValLysThrAenProGluGluLysLysCysPheAspLeuIleSerHisAspArgThr 374
 Db 182 CAGGTGAAGACCAATCCTGAGGAGAGAGAGTGTTCGACCTCATATCATCATGACAGGAG 241
 QY 375 TyrHisPheGlnAlaGluAsePgluGluCysGlnIleTrpMetSerValLeuGlnAen 394
 Db 242 TACCACCTTCCAAAGCGAAGACGAAACAAGATGTGAGATATGGATGTCTGTACTGCAGA 301
 QY 395 SerLysGluGluAlaLeuAenAenAlaPheLysGlyAspAspAenThrGlyGluAenAen 414
 Db 302 AGCAAGAGAGAGCTCTGAACAACGCTTTAAGGGTGATGACACACTGGAGAGAAATAAC 361
 QY 415 IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAenAseP 434
 Db 362 ATCTCCACAGAGCTGACCAAGGAGATCATCTCGAGGTGACAGAGATGACGGCAATGAC 421
 QY 435 ValCysCysAspCysGlyAlaProAseProThrTrpLeuSerThrAenLeuGlyIleLeu 454
 Db 422 GTGTGCTCGCACTGTGGGGACCAAGATCCGACGTGGGTCTCTACCAACCTGGGCATCCTG 481
 QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474
 Db 482 ACTTGATCGAGTGCTCTGGGATCCACGGGAGCTGGGGGTTCATTTCTCAGGATGAG 541
 QY 475 SerLeuThrLeuAsePValLeuGlyThrSerGluLeuLeuLeuAlaLysAenIleGlyAen 494
 Db 542 TCCCTGACATAGATGATTGGGAACGCTGAGCTTCTGCTTGCACAGATATCGGGAT 601
 QY 495 AlaGlyPheAenGluIleMetGluCysCysLeuProAlaGluAsePValLysProAen 514
 Db 602 GCAGGCTTTAATGAATTTATGGAGTGTTCCTACCGCTGAGGACNCAGTCAAGCCCAAC 661
 QY 515 ProGlySerAspMetAenAlaAsePValLysAspTyrIleThrAlaLysTyrIleGluArg 534
 Db 662 CCAGGAGTGCATGATTGCAAGAGAGGACTTACATCACAGCAAGTACATGAGAGAGAGA 721
 QY 535 TyrAlaArgLysLysHisAlaAsePAlaLysLeuHisSerLeuCys 551
 Db 722 TATGACGGAAGAAAGCATGCCGACACCGCGGAGAGCTCCACAGCCTTTC 772

RESULT 6

CO429713

LOCUS

DEFINITION

IMAGE:30686473 5', mRNA sequence.

ACCSSION

CO429713

VERSION

CO429713.1

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 781)

CO429713 781 bp mRNA linear EST 06-JUL-2004
 UI-M-HX0-crw-p-02-0-UI.r1 NIH_BMAP_HX0 Mus musculus cDNA clone

IMAGE:30686473 5', mRNA sequence.

ACCSSION

CO429713

VERSION

CO429713.1

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 781)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
 1..781
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clones="IMAGE:30696473"
 /issue_type="whole eye"
 /dev_stages="newborn (1, 5, 15 days) and embryonic (15, 16, 17, 18 dpc)"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP HX0"
 /note="Organ: Eye; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATAATTACG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
 Pred. No.: 4, 08e-87 Length: 781
 Score: 1351.00 Matches: 251
 Percent Similarity: 98.08% Conservative: 4
 Best Local Similarity: 96.54% Mismatches: 5
 Query Match: 25.62% Indels: 0
 DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x C0429713 (1-781)

Qy 285 SerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHis 304
 Db 2 TCACAACTTCGCCAAAGTACAGCTTACAGTTACATCAACCTCAGGAAACCAAGGAGCAT 61
 Qy 305 GlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLysValTrpGln 324
 Db 62 GGAACAGAGAGAGAGACGGGAACCTCTACAGAGAGAGAGATGGATCCGGAAGTGTGGCAG 121
 Qy 325 LysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArg 344
 Db 122 AAGAGAGAGTGTTCGTTAAATGGCTTCCTCACATCTCCACGGGACATCCGCAACCGG 181
 Qy 345 PropGlnAlaLysLeuAsnLeuThrCysGlnValLysThrAsnProGlnGluLysLys 364
 Db 182 CCACCTGCCAAGCTCAACCTGCTACCTGCCAGGTGAAGACCAATCCTCGAGGAGAAGAAG 241
 Qy 365 CysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGlu 384
 Db 242 TGTTCGACCTCATATCATCATGACAGGAGGTACCATCTTCCAGCGGAGACGAAACAGAA 301
 Qy 385 CysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAlaPhe 404

Db 302 TGTCAAGATATGGATGTCTGTACTGCAGACAGCAAGCAAGAGAGCTCTCTGAACAACGCTTT 361
 Qy 405 LysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIle 424
 Db 362 AAGGGTGATGACAACTGAGGAGAAATAACATCTGCCAAGAGCTGACCAAGAGAGATCATC 421
 Qy 425 SerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspPro 444
 Db 422 TCGAGGTGACAGAGATGACGGCAATGACGTGTGCGACTGTGGGGCACCAGATCCG 481
 Qy 445 ThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArg 464
 Db 482 ACGTGGCTCTTACCAACCTGGGCATCTGACTTGCATCGAGTCTCTGGGATCCACCGG 541
 Qy 465 GluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSer 484
 Db 542 GAGCTGGGGTTCAATTACTCCAGGATGCAGTCCCTGACATTTAGATGTATTGGGAACGTCT 601
 Qy 485 GluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCys 504
 Db 602 GAGCTTCTGCTTGCACCAAGATATCGGAATGCAGCGCTTTAATGAATTTATGGATGTTC 661
 Qy 505 LeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaAArgLysAsp 524
 Db 662 CTACCGTCTGAGGACCCAGCTCAAGCCCAACCCAGCGAGTGACATGATTGCAAGGAAGGAC 721
 Qy 525 TyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAla 544
 Db 722 TACATCAGCAAGTACATGAGGAGGAGATATCAGCGAAAGCATGCCGACACCGCG 781

RESULT 7
 CUS38273
 LOCUS CUS38273 731 bp mRNA linear EST 29-APR-2004
 DEFINITION UI-M-HSO-cqr-k-12-0-UI.r1 NIH BMAP_HSO Mus musculus cDNA clone
 IMAGE:30675611 5', mRNA sequence.
 ACCESSION CUS38273
 VERSION CUS38273.1 GI:46866429
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 731)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pYX-5.
 Location/Qualifiers
 1..731
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clones="IMAGE:30675611"
 /issue_type="Upper Head"
 /dev_stages="embryo 9.5 - 10.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP_HSO"
 /note="Organ: Upper Head; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
Pred. No.: 2,138-81 Length: 731
Score: 1271.00 Matches: 239
Percent Similarity: 99.18% Conservative: 2
Best Local Similarity: 98.33% Mismatches: 2
Query Match: 24.10% Indels: 0
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN538273 (1-731)

Qy 262 GluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGln 281
Db 2 GAACGGGACAGCTGATCACTTCGATTATTTGAAGTCAGCTTCGAGTGAACAG 61
Qy 282 LysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsn 301
Db 62 AAAGAGGACTCAACACTCGCCAAAGTACAGCTTACATCACTCAGGGAAAC 121
Qy 302 LysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLys 321
Db 122 AAGGAGCATGGAACAGAGAGGAGCAACCTCTCAAGAGAGCGATCGGAGAA 181
Qy 322 ValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThr 341
Db 182 GGTGGCAGAGAGAGAGTGTTCCTGTTAAATGCTTCCTCACCATTCTCCACGGCACT 241
Qy 342 AlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGlu 361
Db 242 GCCAACCGGCACCTGCCAAGCTCAACCTGCTAACTGCCAGTGAAGCACTCTGAG 301
Qy 362 GluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAsp 381
Db 302 GAGAAGAAGTGTTCGACCTCATATCATATGACAGAGAGTACCACTTCCAGCGGAAGAC 361
Qy 382 GluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsn 401
Db 362 GAACAGAAGTGCAGATATGGATGTCTGTACTGCGAAGACAGCAAGGAAGAGCTCGAAC 421
Qy 402 AsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLys 421
Db 422 AACGCTTTAAGGTGTATGACAACTGGAGAAATATACATCTCTCAGAGCTGACCAAG 481
Qy 422 GluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysAspCysGlyAla 441
Db 482 GAGATCATCTCGAGGTGCAGAGGATGACGGCAATGACGTGTCTGCGACTGTGGGCA 541
Qy 442 ProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGly 461
Db 542 CAGATCGAGCTGTCTCTACCAACCTGGGCATCTCTGACTGTGATCGAGTCTCTGGG 601
Qy 462 IleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeu 481
Db 602 ATTACCGGGAGCTGGGGTTCATTACTCCAGATGACGTCTCTGACATTAGATGATTG 661
Qy 482 GlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMet 501
Db 662 GGAAGCTCTGAGCTTCTGCTCCCAAGATATCGGGAATGACAGGCTTTAATGAATATG 721
Qy 502 GluCysCys 504
Db 722 GAGTGTTC 730

RESULT 8
CN525450
LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers
1..742
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30638251"
/tissue_type="Upper Head"
/dev_stages="9.5-10.5 dpc"
/lab_host="DHI0B (T1 phage resistant)"
/clone_lib="NIH BMAP HN0"
/note="Organ: Head; Vector: pX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
Pred. No.: 2,138-81 Length: 742
Score: 1271.00 Matches: 242
Percent Similarity: 99.59% Conservative: 3
Best Local Similarity: 98.37% Mismatches: 1
Query Match: 24.10% Indels: 0
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN525450 (1-742)

Qy 164 LysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGlyAlaGluIle 183
Db 3 AAAAGGAGCATGCCAGCTCCAGGATGATTCGTACTGAAATAAGGGGCTGAGATT 62
Qy 184 AlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeu 203
Db 63 GCAGAGGAGATGGAGAGGAAGACGGTTCCTCCAGCTGCAGATGTGTGAGTACCTGCTG 122

CN525450 742 bp mRNA linear EST 29-APR-2004
UI-M-HN0-cnff-f-20-0-UI.r1 NIH_BMAP_HN0 Mus musculus cDNA clone
IMAGE:30638251 5', mRNA sequence.

CN525450
EST.

CN525450.1 GI:46853103

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 742)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mouseeff.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pX-5.

```

204 LysValAsnGluLeuLysLeuLysGlyValAspLeuLeuGlnAsnLeuLeuLysTyr 223
Db 123 AAGGTCATGAATCAAGGTGAAAAGGAGTGATTTGCTTCAGAACTCATAGTAC 182
Qy 224 PheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSerLeuLysPro 243
Db 183 TTTACGCCCAAGTGCATTTTTCAGGATGGATTGAAAGCAGTAGAAGGCTCAAGCCT 242
Qy 244 SerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluArg 263
Db 243 TCCATCGAGAGCGCTCCACGAGACCTCCACACCATCAAAACAGGCCAGGATGAGGAACGG 302
Qy 264 ArgGlnLeuLeuGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGlu 283
Db 303 CGACAGCTGATACAACTTCGAGATATTTTGAAGTCAGCATTTGCAGGTGGAACAGAAAGAG 362
Qy 284 AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGlu 303
Db 363 GACTCACAACTTCGCCAAAGTACAGCCTACAGCTTACATCAACCTCAGGGAACCAAGGAG 422
Qy 304 HisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrp 323
Db 423 CATGAAACAGAGAGGAACGGGAACCTCTACAGAAAGAGCGATGGGATCCGAAAAGTGTGG 482
Qy 324 GlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsn 343
Db 483 CAGAAGAGGAAGTGTTCGTTTAAATTTGGCTTCCTCACCATCTCCACGGCAGCTGCCAAC 542
Qy 344 ArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnProGluGluLys 363
Db 543 CGGCCACCTGCCAAGCTCAACCTGCTAACCCTGCCAGGTGAAGACCAATCTCTGAGGAGAAG 602
Qy 364 LysCysPheAspLeuLysSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383
Db 603 AAGTGTTCGACCTCATATCATACAGAGGAGTACCACTTCCAAAGCGGAAGACGAACA 662
Qy 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403
Db 663 GAATGTCAGATATGGATGTCTGTACTGCAGAACACAGCAAGAAAGAAAGCTCTGNAACGCC 722
Qy 404 PheLysGlyAspAspAsn 409
Db 723 TTAAAGGGTGATGACAAC 740

RESULT 9
BUI72807
LOCUS
DEFINITION AGENCOURT_7966859_NIH_MGC_67_Homo_sapiens_cDNA_clone_IMAGE:6170038
5', mRNA sequence.
ACCESSION BUI72807
KEYWORDS EST.
SOURCE BUI72807.1 GI:22686791
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1536 row: b column: 23
High quality sequence stop: 609.
Location/Qualifiers
1..901

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:6170038"
/tissue_types="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 Kb. Library constructed by Life
Technologies."

ORIGIN
Alignment Scores: 4e-81 Length: 901
Pred. No.: 1269.00 Matches: 257
Score: 1269.00
Percent Similarity: 91.93% Conservat: 5
Best Local Similarity: 90.18% Mismatches: 16
Query Match: 24.07% Indels: 7
DB: Gaps: 2

US-09-914-042-1 (1-1006) x BUI72807 (1-901)
Qy 121 AsnIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 8 TCCGTAATCTCTTCCCTTTGGACAGTTTCTGAAGGGGACCTGAAAGGAGTCAAAGGG 67
Qy 141 AspLeuLysLysProPheAspLeuAlaTyrLysAspTyrGluThrLysIleThrLysIle 160
Db 68 GATCTGAAAAAGCCCTTTTGTATAAAGCTTGGAAAGGACTATGAACAACAAAATAAACCAAGATA 127
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
Db 128 GAAAGAGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATTAAGCGGA 187
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 188 GCGGAATTCGCGAAGAGATGGAAGAGAGAGGCGCTTCTCCAGCTACAGATGTGGAG 247
Qy 201 TyrLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 248 TATCTGCTGAAGTCAACGAAATCAAGATTAAGGAGGAGTAGATTTACTTCAGAACTGTG 307
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240
Db 308 ATCAAAATCTTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAGAGCGTGGAAAGC 367
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 368 CTCAAACTTCCATTAAGACGCTGTCTACGGATCTTCCACAGATCAAAACAGCCAGGAT 427
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLysSerAlaLeuGlnValGlu 280
Db 428 GAAGAAAGAGGAGCTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCAGGTTGAA 487
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 488 CAGAAAGAGGAGTCCCAAAATTCGTGAGAGCACACCTTATAGCTTACATCAGCCTCAGGGA 547
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLysArg 320
Db 548 AACAAAGGACATGGGACCGGAGCGGACCGGACCTCTACAAGAAAGAGTGCAGGATCCGA 607
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 608 ANAGTGTGGCAGAAAAAGGAATGTTCAAGTAAAAATGGTTTTTCTGACCATATCCCATGGT 667
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuThrCysGlnValLysThrAsnPro 360
Db 668 ACCGCTAACCGGCTCTCTGCAAAAGCTCACTGTAACCTGCCAGGGGAAGACCAACCCCT 727
Qy 361 GluGluLysLysCys-PheAspLeu-IleSerHisAspArg-ThrTyrHisPheGlnAla 379

```

Db 728 GAGGAGAAGAGTGTCTTTTGACCTTTATTCACCATGACAGAACTTACCACCTTTCAGG 787
 QY 380 GluAspGlu-----GlnGluCysGlnIle-----TrrMetSerValLeuGlnAsnSer 395
 Db 788 CTGGAAGAGTAAGAACGGAATGTCCAATAATGGGATGCTCTGGCTGGCCAAATTTAG 847
 QY 396 LysGluGlu 398
 Db 848 CCCAAGGAA 956

RESULT 10
 CA328666 784 bp mRNA linear EST 09-JUL-2003
 LOCUS UI-M-F10-cda-o-22-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
 DEFINITION IMAGE:6827087 5', mRNA sequence.

ACCESSION CA328666
 VERSION CA328666
 KEYWORDS EST.

SOURCE GI:24546764
 ORGANISM Mus musculus (house mouse)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1. (bases 1 to 784)
 TITLE NIH-MGC <http://mgc.nci.nih.gov/>
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

FEATURES

source

Location/Qualifiers

1..784
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CS7BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6827087"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_F10"
 /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Rac vector. The library tag
 sequence located between the Not I site and the polyA tail
 is ACCGAGACAG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 5,39e-81 Length: 784
 Score: 1266.00 Matches: 243
 Percent Similarity: 96.15% Conservatve: 7
 Best Local Similarity: 93.46% Mismatches: 10
 Query Match: 24.01% Indels: 0
 DB: 6 Gaps: 0

US-09-914-042-1 (1-1006) x CA328666 (1-784)

QY 508 GluAspSerValLysPrcAsnProGlySerAspMetAenAlaArgLysAspTyrIleThr 527
 Db 2 GAGGACCAGTCAAGCCCAACCCAGGCACTGACATGATTGCAAGGAAGACTACATCACA 61
 QY 528 AlalysTyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeu 547
 Db 62 CCCAAGTACATGAGAGAGAGATATCCACGAAAAAGATGCCGACACCCGCGCGAAGCTC 121
 QY 548 HisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuGluAlaLysTyrAla 567
 Db 122 CACAGCCTTTCCGAGGCGCTCAAGACGAGACATTTTGGGTACTCCAAGCTTATGCT 181
 QY 568 AspGlyValAspLeuThrGluLysIleProLeuAlaAenGlyHisGluProAspGluThr 587
 Db 182 GATGTTGGACCTGACAGAGAAATCCCACTGGCCATGGGCATGAGCCAGATGAGACA 241
 QY 588 AlaleuHisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeu 607
 Db 242 GCCCTCCATCTTGCACTGAGTCACTGTGACCGGACTTCCCTTCACTTGTAGACTTCCTG 301
 QY 608 ValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyr 627
 Db 302 GTCCAGAACAGTGGGAACCTGGATAAACAGACAGCAGGCGAGCAGCCCTGCACCTAC 361
 QY 628 CysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIle 647
 Db 362 TGCTGCTGACAGACAATGCTNAGTGCCTCAAGTCTCTCCGCGGGAAGAGCTCCATT 421
 QY 648 GlulleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLysHisGlu 667
 Db 422 GAGATAGCAATGAGTCAGGAGAGACCCCGTTGGACATTGCTAAGCGCTCAAGCAGCAA 481
 QY 668 HisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisVal 687
 Db 482 CACTGTGAGGAACCTGCTCACTCAGCGCTTGTGCGGCGGTTTAACTCCCATGTTCCAGCTC 541
 QY 688 GluTyrGluTrpArgLeuHisGluAspLeuAspGluSerAspAspAspMetAspGlu 707
 Db 542 GAGTATGTAATGGCGCTGCTGCACGAGACTTGGATGAGAGCAGCAGATGTGGACGAG 601
 QY 708 LysLeuGlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGly 727
 Db 602 AAGCTACAGCCTAGTCTCAACCGGAGGAGACCGGCTGTTAGCTTCTACCAGCTGGGG 661
 QY 728 SerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLys 747
 Db 662 TCCAGCCAGTTTCAGTCCCAATGCTGTGCTTTGGCCAGAGATCTGCACACCTCACCAG 721
 QY 748 GluLysGlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeu 767
 Db 722 GACAAGCAGAGGGGGCTTTGGGCCCCCAGCATCTTGCAGAAATGAGACCTATGGAGCCATCTG 781

RESULT 11

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CN533008 742 bp mRNA linear EST 29-APR-2004
 UI-M-H00-cpx-g-18-0-UI.r1 NIH_BMAP_H00 Mus musculus cDNA clone
 IMAGE:30657089 5', mRNA sequence.

CN533008

EST.

Mus musculus (house mouse)

Mus musculus

1. (bases 1 to 742)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousef1.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. 742
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30657089"
/tissue_type="Upper Head"
/dev_stages="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP H00"
/note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAATGTAAT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

FEATURES
source

ORIGIN

Alignment Scores:
Pred. No.: 1,35e-80 Length: 742
Score: 1260.00 Matches: 243
Percent Similarity: 99.60% Conservative: 3
Best Local Similarity: 98.38% Mismatches: 1
Query Match: 23.90% Indels: 1
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CMS33008 (1-742)

QY 192 ArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluLeuLysLeuLys 211
DB 2 CGGTTCCTCCAGTCGACATGTGTGAGTACCTGCTGAAGTCAATGAATCAAGGTGAAA 61
QY 212 LysGlyValAspLeuLeuGlnAsnLeuLeuLysTyrPheHisAlaGlnCysAsnPhePhe 231
DB 62 AAGGAGTGGATTTGCTTCAGAACTGATCAAGTACTTTTCAGCCCGCATTAATTTTTC 121
QY 232 GlnAspGlyLeuLysAlaValGlnSerLeuLysProSerIleGluThrLeuSerThrAsp 251
DB 122 CAGGATGGATTGAAGACAGTAGAAGCTCAAGCTTCATCGAGACCTTCCACGGAC 181
QY 252 LeuHisThrIleLysGlnAlaGlnAspGluGluArgArgGlnLeuIleGlnLeuArgAsp 271
DB 182 CTCACACCATCAACACAGCCCGAGTGAAGAACGGCGACAGCTGATCACTTCGAGAT 241
QY 272 IleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgLysSerThr 291
DB 242 ATTTTGAAGTCAGCATTCAGGTGGAACAGAAAGAGGACTCACACTTCGCCAAGTACA 301
QY 292 AlaTyrSerLeuHisGlnProGlnGlyValAsnLysGluHisGlyThrGluArgAsnGlySer 311
DB 302 GCCTACACCTTACATCAACCTCAGGGAACACAGGAGCATGGAACAGAGAGACGGGAAC 361
QY 312 LeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLys 331
DB 362 CTCTACAAGAGAGCGATGGGATCCGAAAGTGTGGCAGAGAGAGAAAGTGTTCGGTTAAA 421

QY 332 AsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeu 351
DB 422 AATGGCTTCTTCACCATCTCCACGGCAGCTGCCAACCGCCACTGCCAGCTCAACCTG 481
QY 352 LeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIleSerHis 371
DB 482 CTAACCTGCCAGGTGAAGACCAATCTCTGAGGAGAAGAAGTGTTCGACCTCATATCAT 541
QY 372 AspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerVal 391
DB 542 GACAGGACGTACCACTTCCAAAGCGGAAGACCAAGAATGTTCAGATATGATGTCTGTA 601
QY 392 LeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGly 411
DB 602 CTGCAGAACAGCAAGGAAGAGCTCTGAACAACCCCTTTAAGGGTGTATGACACACATGGA 661
QY 412 GluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThr 431
DB 662 GAAATATAC-ATCGTCCAAAGAGTGACCAAGAGATCATCTCGGAGGTGCAGAGGATGACG 720
QY 432 GlyAsnAspValCysCysAsp 438
DB 721 GGCAATGACGTGTGCTGCGAC 741

RESULT 12

BM944240
LOCUS
DEFINITION
UI-M-EH0p-bvq-c-06-0-UI.r1 NIH_BMAP_EH0p Mus musculus cDNA clone
IMAGE:5695421 5', mRNA sequence.

ACCSSION
BM944240

VERSION
BM944240.1 GI:19427825

KEYWORDS
EST.

SOURCE
Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 740)
NIH-MGC <http://mgs.nci.nih.gov/>.

AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE
Unpublished (1999)

JOURNAL
Contact: Robert Strausberg, Ph.D.

COMMENT
Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. 740
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5695421"
/tissue_type="whole brain"
/dev_stages="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_EH0p"

/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag

sequence located between the Not 1 site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 1,26e-79 Length: 740
Score: 1246.50 Matches: 240
Percent Similarity: 98.78% Conservatives: 3
Best Local Similarity: 97.56% Mismatches: 0
Query Match: 23.64% Indels: 3
DB: 5 Gaps: 1

US-09-914-042-1 (1-1006) x BM944240 (1-740)

Qy 174 IleArgThrGluIleSerGlyAlaGluLeuAlaGluMetGluLysGluArgPhe 193
Db 3 ATTCTACTGAATTAAGCGGCTGAGATTGCAGAGGAGATGGAGAGAAAGCGTTC 62
Qy 194 PheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluLeuLysGly 213
Db 63 TTCAGCTGCGATGTGTGAGTACTCTGTAAGGTCAATGAATCAAGTGAAGAAGGA 122
Qy 214 ValAspLeuLeuGlnAsnLeuLysTyrPheHisAlaGlnCysAsnPheGlnAsp 233
Db 123 GTGGATTTCCTCAGAACTGATCAAGTACTTTTCAGGCCAGTGCATTTTTCAGGAT 182
Qy 234 GlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHis 253
Db 183 GGATTGAAGACAGTAGTAAAGCCTCAAGCCTTCATCGAGAGCCTCTCCAGCGACCTCCAC 242
Qy 254 ThrIleLysGlnAlaGlnAspGluArgGlnLeuLeuGlnLeuArgAspIleLeu 273
Db 243 ACCATCAACAGCGCCAGGATGAGAACCGGACAGCTGATACAACTTCGAGATATTTG 302
Qy 274 LysSerAlaLeuGlnValGluGlnLysGlu-----AspSerGlnIleArgGlnSer 290
Db 303 AAGTCAGATTCAGGTGGAACAGAAAGTCTAGGAGAGACTCAACTTCGCCAAAGT 362
Qy 291 ThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThrGluArgAsnGly 310
Db 363 ACAGCCTACAGCTTACATCAACTCAGGAAACAGGAGCAGTGAACAGAGAGAAACGGG 422
Qy 311 SerLeuTyrLysLysSerAspGlyIleArgLysValTyrGlnLysArgLysCysSerVal 330
Db 423 AACCTCTCAAGAGAGCGGATGGGATCCGGAAGTGTGGCAGAGAGAAAGTGTTCGGTT 482
Qy 331 LysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsn 350
Db 483 AAAATGGCTTCCTCACCATCTCCACCGAAGTGCACACCGGCCACTTCGCNAGTCAAC 542
Qy 351 LeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuSer 370
Db 543 CTGCTAACCTGCAGGTGAAGACCAATCTCGAGGAGAAGAGTGTTCGACCTCATATCA 602
Qy 371 HisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTyrMetSer 390
Db 603 CATGACAGGAGCTACCACTTCACGCGGAAGACGAACAAGATGTCAGATATGGATGTCT 662
Qy 391 ValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThr 410
Db 663 GTACTGCGAAGACAGGAGAGAGAGCTCTGAACAAACGCTTTAAGGGTGTGACAACT 722
Qy 411 GlyGluAsnAsnIleVal 416
Db 723 GGAGAAATAAATCATCGTC 740

RESULT 13
CN537253 738 bp mRNA linear EST 29-APR-2004
LOCUS CN537253

DEFINITION

UI-M-HS0-cqk-j-11-0-UI.r1 NIH BMAP_HS0 Mus musculus cDNA clone

IMAGE:30672898 5', mRNA sequence.

CN537253

VERSION CN537253.1 GI:46865409

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 738)

NIH-MSC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mousefl.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..738

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30672898"

/tissue_type="Upper Head"

/dev_stage="embryo 9.5 - 10.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_HS0"

/note="Organ: Upper Head; Vector: pYX-Asc; Site: 1; Ecor I;

Site 2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with Ecor I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is CGAACTGAAT. This library was created for the University

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
Pred. No.: 6.05e-79 Length: 738
Score: 1237.00 Matches: 235
Percent Similarity: 96.75% Conservatives: 3
Best Local Similarity: 95.53% Mismatches: 7
Query Match: 23.46% Indels: 1
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CN537253 (1-738)

Qy 329 SerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLys 348
Db 3 TCCGTTAAATGGCTTCTCACCATCTCCACGGCACTGCCACCGCCACCTCCCAAG 62
Qy 349 LeuAsnLeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeu 368
Db 63 CTCAACCTGCTTAACCTGCCAGTGAAGACCAATCTGAGGAGAAGAGTGTTCGACCTC 122
Qy 369 IleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTyr 388
Db 123 ATATCATGATGACGAGCGTACCTTCACCGGGAAGACGAACAAGATGTCAGATATGG 182

```

Qy 389 MetSerValLeuGlnAsnSerLysGluAlaLeuAsnAsnAlaPheLysGlyAspAep 408
Db 183 ATGCTGTACTGCAGAACAGCAGAGAAAGCTCTGAACACAGCCTTTAAGGTTGATGAC 242
Qy 409 AsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleSerGluValGln 428
Db 243 AACACTGGAGAAAATAACATCGTCCAGAGCTGACCAAGGAGATCATCTCGGAGGTGCAG 302
Qy 429 ArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTripleuSer 448
Db 303 AGGATGACGGGCAATGACGTGTGCTGCCACTGTGGGGCACCAGATCCGACGTGCTCTCT 362
Qy 449 ThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyVal 468
Db 363 ACCAACTGGGCATCTGACTTGATCGAGTGCTCTGGATCCACCGGAGCTGGGGTT 422
Qy 469 HisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuLeu 488
Db 423 CATTACTCCAGATGTCAGTCCCTGACATTAGATGATTGGGAACGCTCGAGCTTCTGCTT 482
Qy 489 AlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGlu 508
Db 483 GCCAAGAAATATCGGAATGCGAGCTTTAATGAAATTATGGAGTGTGCTTACCGTCTGAG 542
Qy 509 AspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAla 528
Db 543 GACCCAGTCAAGCCCAACCCAGGAGTGCATGATTGCAAGGAGGATCATCATCAGCC 602
Qy 529 LysTyrIleGluArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHis 548
Db 603 AAGTACATGAGAGAGAGATATGCACGGAATAAGCATGCCGACCGCGGCGAAGCTCCAC 662
Qy 549 SerLeuCysGluAlaValLysThrArgAspIlePhe-GlyLeuLeuGlnAlaTyrAlaAs 568
Db 663 AGCCTTTCCGAGGGCGCGTCAAGACACAGACATTTTTTNGGGTTACTCCAAAGCTTATGCTGA 722
Qy 568 pGlyValAspLeuThr 573
Db 723 TGGTGTGACCTGCACA 738

CF539222 737 bp mRNA linear EST 12-SEP-2003
IMAGE:30535336 5', mRNA sequence.
CF539222 CF539222.1 GI:34591591
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 737)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefi.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1.737
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6"
/db_xref="taxon:10090"

```

```

/clones="IMAGE:30535336"
/tissue_type="Whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP GH0"
/notes="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAACTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 8,42e-79 Length: 737
Score: 1235.00 Matches: 238
Percent Similarity: 97.17% Conservative: 2
Best Local Similarity: 96.36% Mismatches: 5
Query Match: 23.42% Indels: 2
DB: 7 Gaps: 0

US-09-914-042-1 (1-1006) x CF539222 (1-737)

Qy 330 VallysAnglyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeu 349
Db 1 GTTAAATAATGGCTTCTCCATCATCTCCACGGCACTGCCAACCGGCCACCTGCCAAGCTC 60
Qy 350 AsnLeuLeuThrCysGlnValLysThrAsnProGluLysLysCysPheAspLeuIle 369
Db 61 AACCTGTACCTGCCAGGTGAAGACCAATCTCGAGGAGAGAGAGTGTTCGACTCATATA 120
Qy 370 SerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTyrMet 389
Db 121 TCACATGACAGGAGCTACCACTTCCAAAGCGGAGACGAACAAGAAATGTTCAGATATGATG 180
Qy 390 SerValLeuGlnAsnSerLysGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsn 409
Db 181 TCTGTACTGCAGAACACGCAAGGAGAGAGCTCTGAACAACGCCCTTTAAAGGGGTGATGACAAC 240
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
Db 241 ACTGGAGAAAATAACATCGTCCAGAGCTGACCAAGGAGAGATCATCTCGGAGGTGCAGAGG 300
Qy 430 MetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTripleuSerThr 449
Db 301 ATGACGGGCAATGACGTGTGCTGCGCATGTGGGGCCACAGATCCGACGTGCTCTCTACC 360
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
Db 361 AACCTGGGCATCTCGACTTGATCGATGCTCTGGGATCCACCGGAGCTGGGGTTCAT 420
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
Db 421 TACTTCCAGGATGCACTCCCTGACATTAGATGATTATGGGAGCTCTGAGCTTCTGCTTGGC 480
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
Db 481 AAGAATATCGGGAATCGAGGCTTTAATGAAATTATGGAGTGTTCCTACCTACCGTCTGAGGAC 540
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
Db 541 CCAGTCAAGGCCAACCCAGGCGAGTGCATGATTTCAGAGGAGGACTCATCATCAGGCCAAG 600
Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549

```

FEATURES
source

```

Db      601  TACATGGAGAGATATGCGGAAAGATGCGACCGCGCGGAAGCTCCACAGC 660
      550  LeuCys-GluAlaValLysThrArgAspIlePhe-GlyLeuLeuGlnAlaValAlaAspG 569
      661  CTTTGGCGAGCGCGTCAAGACAGACAGACATTTTGGGGTTACTTCAAGCTTATGCTGATG 720
      569  lValAspLeuThrGlu 574
      721  GTGTGGACCTGACAGAG 737

```

```

RESULT 15
LOCUS   BU322224
DEFINITION 603488266F1 CSEQCHNG2 Gallus gallus cdna clone CHEST386c18 5', mRNA
sequence.
ACCESSION BU322224
VERSION   BU322224.1 GI:25830225
KEYWORDS EST.
SOURCE   Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 861)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE    A Comprehensive Collection of Chicken cDNAs
JOURNAL  Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE  2235534
PUBMED  12445392
COMMENT  Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

```

```

FEATURES
source
1..861
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST386c18"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN62"
/notes="Organ: heads; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,84e-78 Length: 861
Score: 1229.00 Matches: 236
Percent Similarity: 96.03% Conservative: 6
Best Local Similarity: 93.65% Mismatches: 9
Query Match: 23.31% Indels: 1
DB: 5 Gaps: 0

```

```

US-09-914-042-1 (1-1006) x BU322224 (1-861)
QY      200  GluTyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsn 219
      104  CAGTACCTGCTGAAAGTCAACGAATCAAGATTAAAAAAGGAGTGGAGCTGGCTTCAGAAC 163
QY      220  LeuIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu 239
      164  CTGATCAAGTACTTTTCATGCTCAATGCAATTCTTTCAGGATGGATTAAAAAGCAGTTGAA 223
QY      240  SerLeuLysProSerIleGluThrLeuSerThrPheLeuHisThrIleLysGlnAlaGln 259
      224  AGTCTCAAGCTTCAATGGAGACACTCTCAACAGACCTTTATACGATAAAAACAACACAG 283
QY      260  AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279
      284  GATGAAGAAAGAGACAAATTAAACGAGCTCAGAGATATTTTAAAAATCTGCACCTCCAGTT 343
QY      280  GluGlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGln 299
      344  GATCAGAAAGAGGATTCTCAGATTCTGCAGAGTACAGCTTACAGTTTACATCAGCCTCAG 403
QY      300  GlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIle 319
      404  GGAACACAGGAACATGGCACTGAACGAATGGGAGTCTGTACAAAGAAAGTATGGAATC 463
QY      320  ArgLysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHis 339
      464  AGAAAGTGTGGCAGAAAAGGAGTGTCTCAGTTAAACGGCTTTCTTACAAATTTCCCAT 523
QY      340  GlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsn 359
      524  GGTACAGCTAACCGACCTCCAGCAAGCTCAATCTGTAACTGCTCAAGTGAACAAACAAAC 583
QY      360  ProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAla 379
      584  CCAGAGGAGAAAATGTTTGGACCTCATATCATCATCAGGACATACCATCTTCCAAAGCA 643
QY      380  GluAspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAla 399
      644  GAAGATGACAGGAATGTCAATATGGACATCTGTTCTACAAAACAGCAAGGAGGAAGCT 703
QY      400  LeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeu 419
      704  TTAATAATATGCATTCACAGGAGATGATAACACAGGAGAAAATAATATGCTCCAGGAAGCTG 763
QY      420  ThrLysGluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCys 439
      764  ACAAAAGAGATTATATCTGAAGTCCAAAGGATGACTGGAAATGATGTGTGTGTGACTGT 823
QY      440  -GlyAlaProAspProThrTyrLeuSerThrAsn 450
      824  GGGAGCACAGATTCTACCTGGGTTCACCAANAT 857

```

Search completed: August 4, 2005, 13:57:03
Job time : 7518 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 08:34:52 ; Search time 8647 Seconds
(without alignments)
4749.391 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGAPVPVHFHTAD 1006

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 45554873 seqs, 20411521753 residues

Total number of hits satisfying chosen parameters: 91109746

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0*

Maximum Match 100*

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO spool_h/US09914042/runat_04082005_090634_2150/app query.fasta_1.1159
-DB=Pending Patents NA_Main -OPMT=fastap -SUFFIX=rnppm -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi
-LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09914042 @CGN 1 1 5097 @runat_04082005_090634_2150 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Pending Patents NA_Main:*

- 1: /cgn2_6/ptodata/1/pna/PCTUS1 COMB.seq:*
- 2: /cgn2_6/ptodata/1/pna/PCTUS2 COMB.seq:*
- 3: /cgn2_6/ptodata/1/pna/PCTUS COMB.seq:*
- 4: /cgn2_6/ptodata/1/pna/US06 COMB.seq:*
- 5: /cgn2_6/ptodata/1/pna/US07 COMB.seq:*
- 6: /cgn2_6/ptodata/1/pna/US08 COMB.seq:*
- 7: /cgn2_6/ptodata/1/pna/US081 COMB.seq:*
- 8: /cgn2_6/ptodata/1/pna/US082 COMB.seq:*
- 9: /cgn2_6/ptodata/1/pna/US083 COMB.seq:*
- 10: /cgn2_6/ptodata/1/pna/US084 COMB.seq:*
- 11: /cgn2_6/ptodata/1/pna/US085 COMB.seq:*
- 12: /cgn2_6/ptodata/1/pna/US086 COMB.seq:*
- 13: /cgn2_6/ptodata/1/pna/US087 COMB.seq:*
- 14: /cgn2_6/ptodata/1/pna/US088 COMB.seq:*
- 15: /cgn2_6/ptodata/1/pna/US089 COMB.seq:*
- 16: /cgn2_6/ptodata/1/pna/US090 COMB.seq:*
- 17: /cgn2_6/ptodata/1/pna/US091 COMB.seq:*
- 18: /cgn2_6/ptodata/1/pna/US092 COMB.seq:*
- 19: /cgn2_6/ptodata/1/pna/US092B COMB.seq:*
- 20: /cgn2_6/ptodata/1/pna/US093A COMB.seq:*
- 21: /cgn2_6/ptodata/1/pna/US093B COMB.seq:*
- 22: /cgn2_6/ptodata/1/pna/US094 COMB.seq:*
- 23: /cgn2_6/ptodata/1/pna/US095A COMB.seq:*
- 24: /cgn2_6/ptodata/1/pna/US095B COMB.seq:*
- 25: /cgn2_6/ptodata/1/pna/US095C COMB.seq:*
- 26: /cgn2_6/ptodata/1/pna/US095D COMB.seq:*
- 27: /cgn2_6/ptodata/1/pna/US096A COMB.seq:*
- 28: /cgn2_6/ptodata/1/pna/US096B COMB.seq:*

- 29: /cgn2_6/ptodata/1/pna/US096C COMB.seq:*
- 30: /cgn2_6/ptodata/1/pna/US096D COMB.seq:*
- 31: /cgn2_6/ptodata/1/pna/US096E COMB.seq:*
- 32: /cgn2_6/ptodata/1/pna/US097A COMB.seq:*
- 33: /cgn2_6/ptodata/1/pna/US097B COMB.seq:*
- 34: /cgn2_6/ptodata/1/pna/US097C COMB.seq:*
- 35: /cgn2_6/ptodata/1/pna/US098A COMB.seq:*
- 36: /cgn2_6/ptodata/1/pna/US098B COMB.seq:*
- 37: /cgn2_6/ptodata/1/pna/US098C COMB.seq:*
- 38: /cgn2_6/ptodata/1/pna/US098D COMB.seq:*
- 39: /cgn2_6/ptodata/1/pna/US099A COMB.seq:*
- 40: /cgn2_6/ptodata/1/pna/US099B COMB.seq:*
- 41: /cgn2_6/ptodata/1/pna/US099C COMB.seq:*
- 42: /cgn2_6/ptodata/1/pna/US099D COMB.seq:*
- 43: /cgn2_6/ptodata/1/pna/US099E COMB.seq:*
- 44: /cgn2_6/ptodata/1/pna/US099F COMB.seq:*
- 45: /cgn2_6/ptodata/1/pna/US099G COMB.seq:*
- 46: /cgn2_6/ptodata/1/pna/US100A COMB.seq:*
- 47: /cgn2_6/ptodata/1/pna/US100B COMB.seq:*
- 48: /cgn2_6/ptodata/1/pna/US101A COMB.seq:*
- 49: /cgn2_6/ptodata/1/pna/US101B COMB.seq:*
- 50: /cgn2_6/ptodata/1/pna/US102A COMB.seq:*
- 51: /cgn2_6/ptodata/1/pna/US102B COMB.seq:*
- 52: /cgn2_6/ptodata/1/pna/US103A COMB.seq:*
- 53: /cgn2_6/ptodata/1/pna/US103B COMB.seq:*
- 54: /cgn2_6/ptodata/1/pna/US104A COMB.seq:*
- 55: /cgn2_6/ptodata/1/pna/US104B COMB.seq:*
- 56: /cgn2_6/ptodata/1/pna/US105A COMB.seq:*
- 57: /cgn2_6/ptodata/1/pna/US105B COMB.seq:*
- 58: /cgn2_6/ptodata/1/pna/US106A COMB.seq:*
- 59: /cgn2_6/ptodata/1/pna/US107A COMB.seq:*
- 60: /cgn2_6/ptodata/1/pna/US107B COMB.seq:*
- 61: /cgn2_6/ptodata/1/pna/US107C COMB.seq:*
- 62: /cgn2_6/ptodata/1/pna/US107D COMB.seq:*
- 63: /cgn2_6/ptodata/1/pna/US108A COMB.seq:*
- 64: /cgn2_6/ptodata/1/pna/US108B COMB.seq:*
- 65: /cgn2_6/ptodata/1/pna/US108C COMB.seq:*
- 66: /cgn2_6/ptodata/1/pna/US109B COMB.seq:*
- 67: /cgn2_6/ptodata/1/pna/US109C COMB.seq:*
- 68: /cgn2_6/ptodata/1/pna/US110 COMB.seq:*
- 69: /cgn2_6/ptodata/1/pna/US6000 COMB.seq:*
- 70: /cgn2_6/ptodata/1/pna/US6001 COMB.seq:*
- 71: /cgn2_6/ptodata/1/pna/US6002 COMB.seq:*
- 72: /cgn2_6/ptodata/1/pna/US6003 COMB.seq:*
- 73: /cgn2_6/ptodata/1/pna/US6004 COMB.seq:*
- 74: /cgn2_6/ptodata/1/pna/US6005 COMB.seq:*
- 75: /cgn2_6/ptodata/1/pna/US6006 COMB.seq:*
- 76: /cgn2_6/ptodata/1/pna/US6007 COMB.seq:*
- 77: /cgn2_6/ptodata/1/pna/US6008 COMB.seq:*
- 78: /cgn2_6/ptodata/1/pna/US6009 COMB.seq:*
- 79: /cgn2_6/ptodata/1/pna/US6010 COMB.seq:*
- 80: /cgn2_6/ptodata/1/pna/US6011 COMB.seq:*
- 81: /cgn2_6/ptodata/1/pna/US6012 COMB.seq:*
- 82: /cgn2_6/ptodata/1/pna/US6013 COMB.seq:*
- 83: /cgn2_6/ptodata/1/pna/US6014 COMB.seq:*
- 84: /cgn2_6/ptodata/1/pna/US6015 COMB.seq:*
- 85: /cgn2_6/ptodata/1/pna/US6016 COMB.seq:*
- 86: /cgn2_6/ptodata/1/pna/US6017 COMB.seq:*
- 87: /cgn2_6/ptodata/1/pna/US6018 COMB.seq:*
- 88: /cgn2_6/ptodata/1/pna/US6019 COMB.seq:*
- 89: /cgn2_6/ptodata/1/pna/US6020 COMB.seq:*
- 90: /cgn2_6/ptodata/1/pna/US6021 COMB.seq:*
- 91: /cgn2_6/ptodata/1/pna/US6022 COMB.seq:*
- 92: /cgn2_6/ptodata/1/pna/US6023A COMB.seq:*
- 93: /cgn2_6/ptodata/1/pna/US6023B COMB.seq:*
- 94: /cgn2_6/ptodata/1/pna/US6024 COMB.seq:*
- 95: /cgn2_6/ptodata/1/pna/US6025 COMB.seq:*
- 96: /cgn2_6/ptodata/1/pna/US6026 COMB.seq:*
- 97: /cgn2_6/ptodata/1/pna/US6027 COMB.seq:*
- 98: /cgn2_6/ptodata/1/pna/US6028 COMB.seq:*
- 99: /cgn2_6/ptodata/1/pna/US6029 COMB.seq:*
- 100: /cgn2_6/ptodata/1/pna/US6030 COMB.seq:*
- 101: /cgn2_6/ptodata/1/pna/US6031 COMB.seq:*

102: /cgn2_6/ptodata/1/pna/US6032 COMB.seq.*
 103: /cgn2_6/ptodata/1/pna/US6033 COMB.seq.*
 104: /cgn2_6/ptodata/1/pna/US6034 COMB.seq.*
 105: /cgn2_6/ptodata/1/pna/US6035 COMB.seq.*
 106: /cgn2_6/ptodata/1/pna/US6036 COMB.seq.*
 107: /cgn2_6/ptodata/1/pna/US6037 COMB.seq.*
 108: /cgn2_6/ptodata/1/pna/US6038 COMB.seq.*
 109: /cgn2_6/ptodata/1/pna/US6039 COMB.seq.*
 110: /cgn2_6/ptodata/1/pna/US6040 COMB.seq.*
 111: /cgn2_6/ptodata/1/pna/US6041 COMB.seq.*
 112: /cgn2_6/ptodata/1/pna/US6042 COMB.seq.*
 113: /cgn2_6/ptodata/1/pna/US6043 COMB.seq.*
 114: /cgn2_6/ptodata/1/pna/US6044 COMB.seq.*
 115: /cgn2_6/ptodata/1/pna/US6045 COMB.seq.*
 116: /cgn2_6/ptodata/1/pna/US6046 COMB.seq.*
 117: /cgn2_6/ptodata/1/pna/US6047 COMB.seq.*
 118: /cgn2_6/ptodata/1/pna/US6048 COMB.seq.*
 119: /cgn2_6/ptodata/1/pna/US6049 COMB.seq.*
 120: /cgn2_6/ptodata/1/pna/US6050 COMB.seq.*
 121: /cgn2_6/ptodata/1/pna/US6051 COMB.seq.*
 122: /cgn2_6/ptodata/1/pna/US6052 COMB.seq.*
 123: /cgn2_6/ptodata/1/pna/US6053 COMB.seq.*
 124: /cgn2_6/ptodata/1/pna/US6054 COMB.seq.*
 125: /cgn2_6/ptodata/1/pna/US6055 COMB.seq.*
 126: /cgn2_6/ptodata/1/pna/US6056 COMB.seq.*
 127: /cgn2_6/ptodata/1/pna/US6057 COMB.seq.*
 128: /cgn2_6/ptodata/1/pna/US6058 COMB.seq.*
 129: /cgn2_6/ptodata/1/pna/US6059 COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5253	99.6	5711	41	US-09-949-002-75
2	5253	99.6	5711	41	US-09-949-003C-427
3	5253	99.6	5711	66	US-10-956-157-1144
4	5253	99.6	5711	120	US-60-507-511-1144
5	5253	99.6	5756	34	US-09-770-174-4023
6	5240	99.4	5720	86	US-60-172-360-28138
7	5236	99.3	5810	90	US-60-213-359-1582
8	5236	99.3	5810	90	US-60-213-359-1582
9	5210	98.8	6389	26	US-09-572-411-6530
10	5152	97.7	5670	49	US-10-170-235-8470
11	5146	97.6	5670	115	US-60-452-680-5751
12	4664	88.5	3025	41	US-09-949-002-139
13	4664	88.5	3025	41	US-09-949-003C-925
14	4568.5	86.6	5427	49	US-10-170-235-8660
15	4562.5	86.5	5427	115	US-60-452-680-5750
16	4152.5	78.8	5619	48	US-10-144-771-17053
17	4152.5	78.8	5619	106	US-60-360-207-17053
18	3029	57.4	3775	1	PCT-US02-36151-103
19	3026.5	57.4	6110	62	US-10-796-280-303
20	3026.5	57.4	6110	67	US-10-995-561-189
21	3026.5	57.4	6110	126	US-60-568-219-103
22	3026.5	57.4	6212	62	US-10-796-280-302
23	3026.5	57.4	6212	67	US-10-995-561-188
24	3026.5	57.4	6212	126	US-60-568-219-102
25	3024.5	57.4	3946	1	PCT-US02-36151-96
26	3006	57.0	3974	50	US-10-210-281-73
27	2786	52.8	4304	48	US-10-144-771-9150
28	2786	52.8	4304	106	US-60-360-207-9150
29	2556.5	48.5	4997	2	PCT-US03-30720-325
30	2555.5	48.5	5033	22	US-09-488-725A-1427
31	2555.5	48.5	5033	29	US-09-653-450A-14
32	2555.5	48.5	5033	51	US-10-258-898A-1427
33	2555.5	48.5	5033	51	US-10-286-897-1427
34	2380.5	45.1	5445	102	US-60-324-185-16737
35	2371.5	45.0	5582	28	US-09-625-102-1606

36 2371.5 45.0 5582 47 US-10-062-674-1606
 37 2166.5 41.1 4719 115 US-60-453-050-2647
 38 2166.5 41.1 4719 115 US-60-453-135-2647
 39 2166.5 41.1 4719 116 US-60-466-412-2647
 40 2164.5 41.0 4719 49 US-10-170-235-24226
 41 2089.5 39.6 3812 48 US-10-104-047-981
 42 2089.5 39.6 4050 55 US-10-490-605A-1
 43 2089.5 39.6 4050 55 US-10-490-605A-1
 44 2089.5 39.6 4050 62 US-10-777-860A-38
 45 2044.5 38.8 4256 49 US-10-170-235-28087

ALIGNMENTS

RESULT 1

US-09-949-002-75
 ; Sequence 75, Application US/09949002
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMERISMS IN KNOWN GENES ASSOCIATED
 ; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
 ; TITLE OF INVENTION: AND USES THEREOF
 ; FILE REFERENCE: CL000790
 ; CURRENT APPLICATION NUMBER: US/09/949,002
 ; CURRENT FILING DATE: 2000-01-28
 ; PRIOR APPLICATION NUMBER: 60/231,401
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 10823
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 75
 ; LENGTH: 5711
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-002-75

Alignment Scores:

Pred. No.: 3.37e-217 Length: 5711
 Score: 5253.00 Matches: 1003
 Percent Similarity: 99.70% Conservatives: 0
 Best Local Similarity: 99.70% Mismatches: 3
 Query Match: 99.62% Indels: 0
 DB: 41 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-949-002-75 (1-5711)

Qy 1 MetProaspGlnlleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
 Db 341 ATGCCGACCCAGATCTCCGTGCGAAATTCGTGCCGAGACCCATGAGGACTACAAGCG 400

Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAenThrValAlaAlaTle 40
 Db 401 CCCAGCGCTCCAGCTTCACCCACCGCGCAGTGCCGGAACACTGTGGCGGCATC 460

Qy 41 GluAlaAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
 Db 461 GAGGAGGCTTGGACGTGGACCGATGGTCTTTACAAATGAAGAAATCCGTGAAGCA 520

Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
 Db 521 ATCAACAGCTCTGGCTGGCTCAGTGGAAATGAAGACGAGTACACCGCGCTCTGGAG 580

Qy 81 LysPheGlyGlyAenCysValCysArgAspProPheLeuGlySerAlaPheLeuLys 100
 Db 581 AAGTTTGGCGCAACTGTGTATGCAGATGACCCAGATTTAGGAAGTGCCTTCTTGAAG 640

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120
 Db 641 TTCTCAGTGTTCAAAGGAGTTGACAGCACTTTTCAAAACCTGATTCAAGATATGAAC 700

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
 Db 701 AACATAATCTCTTCCCTTGGACAGTTTGTCTGAAGGGGACCTTGAAGAGGAGTGAAGGG 760

QY 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
DB 761 GATCTGAAAGAGCCCTTTTATGAAAGCTTGGAGGACTATGAAACAAAATAACCAAGATA 820
QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
DB 821 GAAAGGAGAAAGAGAACACGCCAAGCTCCATGGGATGATTTCGACCTGAATAAGCCGA 880
QY 181 AlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
DB 881 GCGGAAATTCGCCGAAGAGATGGAAGAGAGAGCGCTCTTCCACGACACAGATGTGGCAG 940
QY 201 TyrLeuLeuLysValAsnGluLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220
DB 941 TATCTGCTGAAGGTCAAGCAATCAAGATTAAGAAAGGGAGTAGATTTTACTTCAGAATCTG 1000
QY 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240
DB 1001 ATCAATATCTTTCATGCCCAATGCAATTTTTTTCAGGATGGAATCAAGCGGTGGAAGC 1060
QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
DB 1061 CTCRAACCTTCATTGAACGCTGTCTACGATCTTCACAGATCAACAGGCCCCAGGAT 1120
QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
DB 1121 GAAGAAAGAGGACGTTGATACAGCTTCGAGATATTTTGAATCCGCATTTGCAGGTGAA 1180
QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
DB 1181 CAGAAAGAGGACTCCCAATTCGTACAGACACAGCTTATAGCTTACATCAGGCTCAGGGA 1240
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
DB 1241 AACAGGAACATGGACCGGACCGGACCGCTCTACAGAAAGAGTGACGGGATCCGA 1300
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
DB 1301 AAAGTGTGCGCAAAAGGAAATGTTTCAGTTAAATAATGGTTTTCTGACCATATCCCATGT 1360
QY 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
DB 1361 ACCGCTAACCGGCTCTGCAAGCTCAACCTGCTAACCTGCGAGTGAAGACCAACCT 1420
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
DB 1421 GAGGAGAGAGAGTGTCTTGACCTTATTTACATGACAGACTTACCATCTTCAAGCTGAA 1480
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluAlaLeu 400
DB 1481 GATGAACAGGAATGTCAATATGGATGTCTGTGCTGCAAAATAGCAAAAGAGAAGCTTTA 1540
QY 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
DB 1541 AACAAATGCAATTAAGGGGATGACAATATCTGGAGAAATAACATCGTCCAAGAACTGACA 1600
QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
DB 1601 AAGGAGATCACTCAGAGTGCAGAGGATGACGGCAATGACCTCTGTGACTGTGGG 1660
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
DB 1661 GCGCAGATCCTTACATGGCTTTCCACCAACCTGGGCATCTGACCTGCATCGAGTGTTC 1720
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
DB 1721 GGAATCCACCGAGAGCTGGGGTTCATTATTCAGGATGCAGCTCCGCTGACCTTAGATGTA 1780
QY 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
DB 1781 CTGGGAACATCTGAGCTGCTGCTCCCAAGAAATTTGGAAATGACAGGCTTTAATGAGATC 1840
QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520

DB 1841 ATGGAATGTTCCTTACCAGCTGAGACTCAGTCAAAACCAACCCAGGAGGACATGAA 1900
QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
DB 1901 GCAAGAAAGGACTACATCAGACCCAGTACATCAGAGGAGATACGCAAGGAAGAGCAC 1960
QY 541 AlaAspAsnAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
DB 1961 GCGGATAACGCGGCGAAGCTTCACAGTCTTTGCGAGCGCTCAAAACAGAGATATTTTT 2020
QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
DB 2021 GGATGTCTCCAGCTTATGCTGATGGTGTGATCTTACGGAATAAATCCCACTGGCCAAC 2080
QY 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
DB 2081 GGACATGACCGGATGAACCGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2140
QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
DB 2141 CTTCACTGTAGACTTTTTTAGTTTCAAGACAGTGGAACTTGGATAAACACAGACAGGAAA 2200
QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
DB 2201 GGCAGCAGCCCTGCACTACTGCTGCTGACCCGACATGCGAGTGCCTCAAGTTGCTC 2260
QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
DB 2261 CTGGGGGGAAGGCTTCATCGAGTACCAACAGTCCAGGAGACTCCGCTGGACATT 2320
QY 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg 680
DB 2321 GCCAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGTGACCCAGCCTTATCTGGAAGA 2380
QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
DB 2381 TTTAATCTCAGTTCACTGTAATATGAAATGGCGACTACTCCACGAAGACCTTGATGAA 2440
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
DB 2441 AGTGAATGACATGATGATGAGAAATTCAGGCCAGTCCCAACCGCGGGAGACCGGCC 2500
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
DB 2501 ATCAGCTTCTACCAGCTGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
DB 2561 GATGCTGCAACCTTGCACAGGAGAGCAGAGGCTTTTCATGCCAGCATCTTGACAGAT 2620
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
DB 2621 GAGACTTACGAGAGCCCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGACGCCCCAGC 2680
QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
DB 2681 ACCACCGCGCCCGCCCTTCTCCACGGAATGTTGGCAAAAGTTTCAGACAGCCTCTCT 2740
QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyLysArgGlnArgSer 820
DB 2741 GCTAACACCTTGTGGAAGACAAACTCTGTAAGTGTGACGGTGGGAAGCGCGAGCATCT 2800
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
DB 2801 TCGTCAGATCCGCGAGCTGTCTCCACCGCTGCCCTCTTTCGGGTGACATCTACCANT 2860
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
DB 2861 CCCCTGACCCCGCGCGCCCGCCCGCTGCGCAAGACGCCAGCGTAAATGGAGCCCTTG 2920
QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880

```
Db 2921 AGCCAGCCGAGCAAGCCTGCCCGCTGGGATCTCACAGATCAGGCCCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCCGCCAGCGCTCCCGCAGAGAAGCCTGCCCGGGGGCTGCACAAGTCCACC 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACTGACCAACAAAGGCCCAACGAGAGAGACTGTGGATCTCTGCAACGGAAGCTCTG 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCAATGCTATGCTGCTGAGCCCTGACGCCCTGCACCATGCTAGGAAGTCG 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCCTTAAGCGGTGAAGCGCTCTATAACTGTGTGGCTGCACAACCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValLlelleValAspGlyGluGluAspGlnGlu 980
Db 3221 CATGAGCTCACTCTCCAGGGGATGTGATCATCTGTGGACGGGGAGAGACCGAG 3280
Qy 981 TrpTrpLleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATTGGCCACATTGATGAGATCTCTGTGCGCAAGGCGCATTCCTCGGTGTCTTT 3340
Qy 1001 ValHisPheLleAlaAsp 1006
Db 3341 GTGCACCTTTATCGCTGAC 3358

RESULT 2
US-09-949-003C-427
; Sequence 427, Application US/09949003C
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPYCHIAIRIC
; TITLE OF INVENTION: DISORDERS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000791
; CURRENT APPLICATION NUMBER: US/09/949,003C
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,446
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 74065
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 427
; LENGTH: 5711
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-949-003C-427

Alignment Scores:
Pred. No.: 3,37e-217 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 41 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-949-003C-427 (1-5711)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCCGACCAAGATCTCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTACNAGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCAGCGCTCCAGCTTCAACACCCCGCAGCGCAGTCCGCGAACACTGTGGCGGCCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 461 GAGGAGGCTTTGGACGTGGACGGATGTTCTTTTACAAAATGAAGAAATCCGTGAAAGCA 520
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
```

```
Db 521 ATCAACAGCTCTGGCTGGCTCAGCTGGAATAAGAGCAGTACACCCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCACAGATTTAGGAAGTGCCTTCTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTTACAAAGGAGTTGACAGCACTTTTCAAAAACCTGATTCAGAAATATGAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATATATCTCTTCCCTTGGACAGTTTCTGAAGGGGACCTGAAAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysLleThrLysLle 160
Db 761 GATCTGAAAAAGCCTTTTGTATAAAGCTTGAAGGACTATGAACAACAAAATAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLleSerGly 180
Db 821 GAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATTAAGCGGA 880
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTCGCGAAGAGATGGAAGAGAGAGGGCGCTTCTCCAGCTACAGATGTGCGAG 940
Qy 201 TyrLeuLysValAsnGluLleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 941 TATCTGTGAAGTCAACGAAATCAAGATTAAGGAGGAGTAGATTTACTTTCAGAACTCG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATACTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAGGCGCTGGAAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTTCATTTGAACGCTGTCTACGGATCTTCACAGCATCAACAGGCCCCAGGAT 1120
Qy 261 GluGluArgArgGlnLeuLleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGCGAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCGAGTTGAA 1180
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGA 1240
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAAAGGACATGGGACCGAGCGGAACGGCAGCCTCTACAAGAGAGTGCAGCGGATCCGA 1300
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAAGTGTGGCAGAAAAGAAATGTTTCAAGTAAAAAATGTTTTCTGACCATATCCCATGGT 1360
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAGAGCTCAACCTGTCTAACCTGCCAGGTGAAGCAACCCCT 1420
Qy 361 GluLysLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAAGAGTCTTTGACCTTATTTTACATCAGAGACTTACCACCTTTCAAGCTGAA 1480
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATCAACAGGAATGTCAAAATATGATGTCTGTGTCGCAAAAATAGCAAAAGAAAGCTTTA 1540
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACAAATGATTTAAGGGGATGACAACTCTGGAGAAATACATCTCGTCCCAAGAACTGACA 1600
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
```

Db 1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGTCGTGTCGACTGTGGG 1660
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCAGATCCTACATGGCTTTCCACCAACCTGGGCATCCTGACCTGCATCGAGTGTTC 1720
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrIleuAspVal 480
Db 1721 GGAATCCACCCAGAGCTGGGGGTTTCATTTATTCAGGATGCAGTCCCTGACCTTAGATGA 1780
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGGAACATCTGAGCTGCTGCTGCCCAAGATATTGGGAATGCAAGGCTTTAATGAGATC 1840
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATGTTGCTACCAAGCTGAGACTCAGTCAAAACCCACCCAGGCGAGCATGAAT 1900
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGAACAC 1960
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 GCGGATACGCGGGAAGCTTCACAGTCTTTGCGAGGCCGTCAAAACAGAGATATTTTT 2020
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATTGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAAATAATCCCACTGGCCAA 2080
Qy 581 GlyHisGluProAspGluThrAlaLeuHisIleuAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCGGATGAACCGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2140
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2141 CTTACATGTTAGACTTTTTAGTTAGAACAGTGGAAACCTGGGATAAACACAGAGGAAA 2200
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2201 GGCAGCAGCCCTGCATCTACTGCTGCTGACCGACATGCGAGTGCTCAAGTTGCTC 2260
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2261 CTGCGGGGAAGGCTCTCCATCGAGATGCAAAACGAGTCAGGAGACTCCCGCTGGACATT 2320
Qy 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCCTCAAGACGACGACCTGTGAGAGCTGTGACCCCAAGCCTTATCTGGAAGA 2380
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTAATTTCTACGTTACGTTGATATGAATGGCGACTACTCCAGGAAGACTGTGATGAA 2440
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2441 AGTCATCAGCATGATGATGAGAAATTCAGCCCAAGTCCCAACCGCGGGAAGACCGGCC 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACCAAGCTGGGCTCAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAACCTTGCCAAAGAGAGAGAGAGGGCTTTTCATGCCAGCATCTTGCAGAA 2620
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2621 GAGACTTACGAGAGCCCTCTCAGTGGCAGCCACCTCCCGCCAGCTGCAGCCCCCAGC 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACGAGCGCCCCCGCTTCTCCACGGAATGTTGGCAAAAGTTTCAGACAGCCTCTCT 2740

Qy 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyLysSerArgGlnArgSer 820
Db 2741 GCTAACACCTGTGGAGACAAACTCTGTAGTGTGGAGCGGTGGGAAGCCGCGAGCATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCGAGCTGTCCATCCACCGTGCCTCTTTCGCGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCAAGCGCCGCCACCTGTTCCCAAGAGCCCGAGGCTTAATGGAGCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGAGCAAGCCTGCCCGCTGGGATCTCACAGATCAGGCCCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCAAGCGCCTCCCGCAGAAAGAGCTGCGCGGGGGCTGACAAAGTCCACC 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCATGACCAACAAGGCCAACCGAGAGGACTGTGGATCTCTCTGCAACCGAAGCTCTG 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGCTCAATGCTATGCTCTGTCAGCCCCCTGCACCCATCGCTAGGAAGTCGAG 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCCTCAAGCGGTGAAGCGCTCTATAACTGTGTGGCTGACAACCCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACTTCTTCGAGGGGATGTGATCATCTGTGGACGGGAGAGACCGAGAG 3280
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATTGGCCACATTTGATGGAGATCTCTGTCGCAAGGCGCATTCCTCCGCTGTCT 3340
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3341 GTGCACCTTTATCGCTGAC 3358
RESULT 3
US-10-956-157-1144
; Sequence 1144, Application US/10956157
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1144
; LENGTH: 5711
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-1144
Alignment Scores:
Pred. No.: 3,37e-217 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 66 Gaps: 0
US-09-914-042-1 (1-1006) x US-10-956-157-1144 (1-5711)

Qy 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleAla 20
Db 341 ATGCCGACCAAGATCTCCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTCAAGAGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCAGGCTCCAGCTTCAACCCCGCACGGCGGCGAGTCCCGGACACTGTGGCGGCCATC 460
Qy 41 GluGluAlaLeuAppValAspArgMetValIleuTyrIysMetLysLysSerValIysAla 60
Db 461 GAGGAGGCTTTGGACCTGCACCGGATGTTCTTTACAAAATCAAGAAATCCGTGAAGCA 520
Qy 61 IleAsnSerSerGlyValAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGGCTCGGCTCACGTGGAAATGAAGAGCAGTACCCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProaspLeuGlySerAlaPheLeuIys 100
Db 581 AAGTTTGGCGCACTGTGTATGCAGAGATGACCCAGATTTAGGAGTGCCTTCTTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAAACCTGATTTCAGAAATGAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTGAAGGGGACCTGAAAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 761 GATCTGAANAAGCCTTTTGATTAAGCTTGGAGGACTATGAACCAAAATTAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTGGGACTGAAATAAGCGGA 880
Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTTGGCGAAGAGATGGAAAGGAGAGGCGCTCTTCCAGCTACAGATGTGCGAG 940
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 941 TATCTGCTGAAGTCAACGAAATCAAGATTAAAGAGGAGTAGATTACTTCAGATCTG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATCTTTCATGCCCAATGCAATTTTTTTCAGGATGGACTCAAAAGCCGTGAAAGC 1060
Qy 241 LeuLysProSerIleGluThrIleuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACCTTCATTGAACCGCTGCTACGGAATCTTCACAGATCAAAACAGGCCCCAGGAT 1120
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGNAAGAGGCGAGTTCATACAGCTTCGAGNATATTTGAAATCCGCATTCGAGGTGAA 1180
Qy 281 GlnLysGluAppSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAATTCGTCAAGACACAGCTTATAGCTTACATCAGCCTCAGGGA 1240
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAGGAACATGGGACCGAGCGGAACGGCGACCTCTCAAGAAAGAGTGACGGGATCCGA 1300
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAAGTGTGGCAGAAAGGAATGTTTCAGTTAAAAATGTTTCTGACCCATATCCCATGGT 1360
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGCTTAACTGCTGAGGAGTGAAGACCAACCT 1420

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAAGAGTGTCTTTGACCTTATTTTCATGACAGACTTACCACCTTTCAGAGCTGAA 1480
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATCAACAGGAATGTCAAAATATGATGTCTGTGCTGCACAAATAGCAAAAGAAAGCTTTA 1540
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACAAATGCATTTAAGGGGATGACAATCTGGAGAAAATAACATCGTCCCAAGAACTGACA 1600
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGGAGATCATCTCAGAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGGG 1660
Qy 441 AlaProaspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCCAGATCCTACATGGCTTTCCACCAACCTGGGCATCTCTGACCTGCATCGAGTGTTC 1720
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1721 GGAATCCACCGAGAGCTGGGGGTTTCATTTATCCAGGATGCACTCCCTGACCTTAGATGTA 1780
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGGAACATCTGAGCTGTCTGCCAGNATATTGGGAATGCAAGGCTTTAATGAGATC 1840
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAACCCAGGCGACATGAAT 1900
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAGGACTACATCACAGCCCAAGTACATCCAGAGGAGATACGCAAGAAAGAACAC 1960
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 GCGGATTAACGGCGGAGAGCTTCACAGCTTTTGGCGGCCCTCAAAACGAGAGATATTTT 2020
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATTGCTCCAAGCTTATGCTGTGATGCTTACGCGAATAAATCCCACTGGCCCAAC 2080
Qy 581 GlyHisGluProaspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCGGATGAACCGCCCTCACCTTCAGTCAGATCCGTGGATCGAACCTCT 2140
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAppLysGlnThrGlyLys 620
Db 2141 CTTACATTTGTAGACTTTTGTAGTTTCAGNACAGTGGGAACCTTGGATTAACACAGAGGAAA 2200
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2201 GGCAGCAGCGCTGCCTACTGCTGCTGACCCGACAATGCGAGTGCCTCAAGTTGCTC 2260
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2261 CTGGGGGGAAGGCTTCATCGATAGCAACAGAGTCAGAGAGACTCCCGCTGGACATT 2320
Qy 661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCTCAAGCACGAGCACTGTGAGAGGCTGCTGACCCCAAGCCTTATCTGGAAGA 2380
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAppLeuAspGlu 700
Db 2381 TTTAATTCTCAGTTACGTTGAATATGAATGGCGACTACTCCAGAAAGACCTCGATGAA 2440
Qy 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2441 AGTATGACGATGGATGAGAAATTGCAGCCCAAGTCCCAACCGGCGGGAAGACCGGCC 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740


```
Db 2501 ATCAGCTCTTACCAGCTGGCTCCACACAGCTTTCAGTCTAAGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAen 760
Db 2561 GATGCTGCAACCTTGGCAGGAGGAGCAGAGGCTTTCATGCCAGCATCTTGGCAGAT 2620
Qy 761 GluThrTyGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAAsnProSer 780
Db 2621 GAGACTTACGAGGAGCTCTGAGTGGCAGCCACCTCCGCGCAGCTGACGCTGAGCCAGC 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAenValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACAGGCGCCCGCTTCTCCACGGAATGTTGGCAAGTTTACACAGCTCTCTCT 2740
Qy 801 AlaAenThrLeuTrpLysThrAenSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTTGTGGAGACCAACTCTGTAGTGTGGAGCGTGGAGCGCGCAGCATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAen 840
Db 2801 TCGTCAGATCGCAGCTGTCCATCCACCGCTGCCCTCTTTCGGCTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGCAGCCCGCCACCCGTTGCCAAGACGCCCGCAGCGTAATGGAAGCCTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGCAGCAAGCTGCCCGCTGGGATCTCAGAGTCAGCCGCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCGCAGCGCTCCCGCAGAGAAGCCCTGCCCGGGGCTGACAAGTCCACC 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACGTACCAACAAAGGCGCAACCGAGAGGAGCCGTGTGGATCTCTCTGCAACGGAAGCTGT 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCATGCTATGCTGTCTGAGCCCGCTGACCCATCCCTAGGAGTGGCAG 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyAsnCysValAlaAspAenPro 960
Db 3161 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATTAAGTGTGCTGACAACCCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValLleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTCGACGGGAGGAGCAGGAG 3280
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATGGCCACATGATGGAGATCCTGTGTCGCAAGGCGCATTTCCCGTGTCAATT 3340
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3341 GTGCACTTTATCGCTGAC 3358
```

RESULT 4

```
US-60-507-511-1144
; Sequence 1144, Application US/60507511
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: AM 101081
; CURRENT APPLICATION NUMBER: US/60/507,511
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 203623
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1144
```

```
; LENGTH: 5711
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-507-511-1144
Alignment Scores:
Pred. No.: 3,376-217 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservat: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
Gaps: 120
```

US-09-914-042-1 (1-1006) x US-60-507-511-1144 (1-5711)

```
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrlsAla 20
Db 341 ATGCGGACCAAGATCTCGTGTGGATTCGTGGCCGAGACCCATGAGGACTACAGGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAenThrValAlaAla 40
Db 401 CCCACGGCTCCAGCTTCCACACCGCAGCGGAGTGCAGAACACCTGTGGCGCCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrlsMetLysLysSerValysAla 60
Db 461 GAGGAGGCTTGGACGTGGACCGGATGTTCTTTACAAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAenSerSerGlyLeuAlaHisValGluAenGluGlnTyrlsThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGCTCAGCTGGAAATGAGAGCAGTACACCGAGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATCACCCAGATTTAGGAAGTGCCTTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAenLeuIleGlnAenMetAen 120
Db 641 TTCTCAGTGTTTACAAAGAGAGTTGACAGCAGCTTTTCAAAAACCTGATTCAGAATATGAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValysGly 140
Db 701 AACATAATCTCTTCTCCCTTTGGACAGTTTCTGAAAGGGGAGCTGAAAGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrlsThrLysIleThrLysIle 160
Db 761 GATCTGAAAAGCCCTTTTGATAAGCTTGGAGGAGCTATGAACAAAAATAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
Db 821 GAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATAAGCGGA 880
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTTGCGAAGAGATGGAAGAGGAGGCGCTTCTCCAGCTACAGATGTGCGAG 940
Qy 201 TyrLeuLeuLysValAenGluIleLysIleLysLysGlyValAspLeuGlnAenLeu 220
Db 941 TATCTGTGAAGTCAACGAAATCAAGATTTAAAGGCGAGTAGATTTACTTCAGAACTCTG 1000
Qy 221 IleLysTyrlsPheHisAlaGlnCysAenPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATCTTTCATGCGCAATGCAATTTTTTTCAGGATGGACTCAAGCGCTGGAAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAen 260
Db 1061 CTCAACCTTCCATGGAACCGCTGTCTACGAGATCTTTCACAGATCAACAGCGCCAGGAT 1120
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGNAGAGGAGGAGTGTGATACAGCTTCGATATTTTGAATTCGCGATTCGAGGTGAA 1180
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrlsSerLeuHisGlnProGlnGly 300
```

1181 CAGAAAGGAGACTCCCAAAATTCGTGAGAGCAGAGCTTATAGCTTACATCAGCCTCAGGGA 1240
Db
301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Qy
1241 AACAAAGGAACATGGACCGGACCGGACCGCTCTACAAAGAGAGTGAGCGGATCCGA 1300
Db
321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Qy
1301 AAAGTGTGGCAGAAAGGAATGTTCAAGTAAATAGTTTCTGACCATATCCCATGGT 1360
Db
341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Qy
1361 ACCGTAACCGGCTCTCTGCAAGAGCTCAACCTGTCTAACCTGCCAGGTGAAGACCAACCT 1420
Db
361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Qy
1421 GAGGAGAAGAAGTGTCTTGGACCTTATTTACATGACAGAATCTTACCACCTTTCAAGCTGAA 1480
Db
381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Qy
1481 GATGACAGGAATGTCAAAATATGGATGTCGTGTCGCAAAATAGCAAAAGAGAGCTTTA 1540
Db
401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGlnLeuThr 420
Qy
1541 AACAAATGCATTTAAGGGGATGACAAATACTGGAGAAAATAACATCGTCCAAAGACTGACA 1600
Db
421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Qy
1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGTCTGCTGTGACTGTGGG 1660
Db
441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Qy
1661 CGCCAGATCTCATATGCTTTCACCAACCTGGGCATCTTGACCTGCATCGATGTGCTCC 1720
Db
461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Qy
1721 GGAATCCACCGAGAGCTGGGGTTCATTATTCAGGATGCAGTCCCTGCACCTTAGATGTA 1780
Db
481 LeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Qy
1781 CTGGGAACATCTGAGCTCTGCTGCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC 1840
Db
501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Qy
1841 ATGGNAATGTTGCTTACCAAGCTGAGACTCAGTCAAAACCCCAACCCAGGCGAGCATGAA 1900
Db
521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Qy
1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGACAC 1960
Db
541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Qy
1961 CGCGATACCGGCGGAAGCTTACAGTCTTTCGAGAGCGCGTCAAAACAGAGAGATATTTT 2020
Db
561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Qy
2021 GGATTCCTCAAGCTTATGCTGATGGTGTGGATCTTACGGAATAAATCCCACTGGCCAA 2080
Db
581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Qy
2081 GGACATGAGCGGATGAACCGCCCTCCACCTTGCAGTCAGTCCGTGGATCGAACCTCT 2140
Db
601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrClyLys 620
Qy
2141 CTTCACTTGTAGACTTTTATAGTTCAGAACACAGTGGGAACCTGGATATAACAGACGGGAAA 2200
Db
621 GlySerThrAlaLeuHisTyrCysValLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Qy
2201 GGCAGCAGCCCTGCATCTACTGCTGCTCCACCGCAATGCGAGTGTCTCAAGTTGCTC 2260
Db
641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Qy
2261 CTGGGGGGAAGGCCCTCCATCGAGATACAAACGAGTCAGGAGACTCCGCTGGACATT 2320
Db

Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCTCAAGCAGCAGCAGCTGTGAGAGCTGTGACCAAGCCTTATCTGGAAGA 2380
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTTAAATTCACGTTACGTTGAATATGAATGGCGACTACTCCACGAAGACCTGGATGA 2440
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2441 AGTGATCAGACATGATGAGAAATTCAGCCAGTCCCAACCGCGGGAAGACCGGCC 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACCAAGCTGGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAACCTTGCNAGGAGAGAGAGGGCTTTTCATGCCAGCATCTTCAGAAAT 2620
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaIleProSer 780
Db 2621 GAGACTTACGAGCGCTCTCAGTGGCAGGCCACTCCCGCCAGCCTGCAGCGCCCGCAGC 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACCGCGCCCCCGCTTCTCCACGGAATGTGGCAAGTTCAGACAGCCTCTCTCT 2740
Qy 801 AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACCCCTGTGGAAGACAAACTCTGTAAGTGTGGACGGTGGAGCGCGCAGCATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCGAGCTGTCCATCCCGCTGCCCTCTTCGCGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCGTGACCCCGCCCGCCCCCAGCGCTTCCCAAGAGCGCCAGCGTAATGGAAGCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGGAGCAAGCTGCCCGCTGGGATCTCAGACATCAGCGCCCGCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCGCCAGCGCTCCCGCAGAAGAAGCTTCGCGCGGGGCTGACAAGTCCACC 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCATGTACCAACAAGGCGCAACCGAGAGGAGCTGTGGATCTCTCTGCNACCGAAGCTCTG 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCCTCTGTCCATGTATGCTGTGCTGCGAGCGCTTCAGCCCGCTTCAGGAAGTCCGAG 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCGCTAAGCGGTGAAGCGCTCTATAACTGTGTGTGGCTGACAACCCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGTGAGCGGGAGGAGGACCGAGAG 3280
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATGGCCACATTTGATGGAGATCTCTGTCGAAAGGCGCATTTCCCGGTGTCAATT 3340
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3341 GTGCACCTTTATCGCTGAC 3358

RESULT 5

US-09-174-174-4023
 ; Sequence 4023, Application US/09770174
 ; GENERAL INFORMATION:
 ; APPLICANT: Shyjan, Andrew W.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
 ; TITLE OF INVENTION: THEREFOR
 ; FILE REFERENCE: 1600.2057-001
 ; CURRENT APPLICATION NUMBER: US/09/770,174
 ; CURRENT FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: US 60/178,835
 ; PRIOR FILING DATE: 2000-01-28
 ; NUMBER OF SEQ ID NOS: 4376
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4023
 ; LENGTH: 5756
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-174-174-4023

Alignment Scores:

Pred. No.: 3.39e-217 Length: 5756
 Score: 5253.00 Matches: 1003
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 3
 Query Match: 99.62% Indels: 0
 DB: 34 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-770-174-4023 (1-5756)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleVala 20
 Db 334 ATGCGGACCAAGATCTCGGTGCGAATTCGTGGCCGAGACCCATGAGGACTTACAAGCG 393
 Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 Db 394 CCCAGGCTTCAGCTTCACCAACCGGACGCGGAGTCCCGAAGACACTGTGGCGCCATC 453
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrIleMetLysSerValIleVala 60
 Db 454 GAGGAGGCTTTGGAGCTGACCGGATGCTTCTTACAAATGAAGAAATCCGTGAAGCA 513
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnThrGlnAlaLeuGlu 80
 Db 514 ATCAACAGCTCTGGCTGGCTCAGTGGAAATGAAGAGCAGTACACCGAGCTCTGGAG 573
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
 Db 574 AAGTTTGGCGCACTGTGTATGCAGATGACCCAGATTTAGGAAGTGCCTCTGAAG 633
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
 Db 634 TTCTCAGTGTACAAAGAGTGTGACACACTTTCAAAAACCTGATTCAGATATGAAC 693
 Qy 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIleGly 140
 Db 694 AACATAATCTCTCTCCCTTGGACAGTTTGTGAAGGGGAGCCTGAAAGAGTGAAGGG 753
 Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysIleThrLysIle 160
 Db 754 GATCTGAAAAGCCCTTTTGTAAAGCTTGAAGGAGTATGAACAAAATTAACCAAGATA 813
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
 Db 814 GAAAGGAGAAAAGGAACACGCCAAGCTCCATGGGATGATTCGAGCTGAATTAAGCGGA 873
 Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
 Db 874 GCGGAAATTCGGAAGAGATGAAAGGAGAGGCGCTCTTCCAGCTACAGATGTGCGAG 933
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
 Db 934 TATCTGCTGAAGGTCACGAAATCAAGATTAAAAAGGGAGTAGATTTTACTTTTCAAGATCTG 993

Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
 Db 994 ATCAAAATCTTTTCATGCGCAATGCAATTTTTTTCAGGATGAGCTCAAAAGCCGTGAAAGC 1053
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
 Db 1054 CTCAAACTTCCATGAAACGCTGTCTACGAGTCTTTCACAGATCAAAAGCCGAGGAT 1113
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
 Db 1114 GAAGAAAGAGCAGCTGATACAGCTTCAGATATTTTGAATCCGCATTCGAGGTTGAA 1173
 Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
 Db 1174 CAGAAAGAGGACTCCAAATTCGTACAGACAGCTTATAGCTTACATCAGCCTCAGGGA 1233
 Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArg 320
 Db 1234 AACAAAGAACATGGAGCCGAGCGGACCTCTTCAAGAGAGTGAAGCGGATCCGA 1293
 Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
 Db 1294 AAGTGTGGCAGAAAGGAAATGTTTCAGTTAAATGTTTTTCTGACCATATCCCATGT 1353
 Qy 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
 Db 1354 ACCGTAACCGGCTCTCTGCAAGCTCAACTGCTAACCTGCGAGGTGAAGACCAACCT 1413
 Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
 Db 1414 GAGGAGAAAGATGCTTTGACCTTATTTTCACATGACAGAACTTACCACTTTTCAAGCTGAA 1473
 Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
 Db 1474 GATGACAGGAAATGTCAAAATATGATGCTGTGTGCAAAATAGCAAGAGAGAGCTTTA 1533
 Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
 Db 1534 AACAAATGCAATTAAGGGGATGACAACTACGAGAAATAACATCGTCCAAGAACTGACA 1593
 Qy 421 LysGluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
 Db 1594 AAGAGATCACTCTCAGAAAGTCAGAGGATGACGGCAATGACGTCTGTGTACTGTGGG 1653
 Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
 Db 1654 GCGCAGATCTTACATGCTTTCCCAACCTGGGCATCTCTGACCTGATCGAGTGTCC 1713
 Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
 Db 1714 GGAATCCACGAGAGCTGGGGGTTTCATTTATCCAGGATGACGTCTCTGACCTTAGATGTA 1773
 Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
 Db 1774 CTGGGAACATCTGAGCTGCTCTGCCAAGAAATATGGAAATGCAAGGCTTAAATGAGATC 1833
 Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
 Db 1834 ATGAATGTTCCTTACCGAGTGAAGCTCAGTCAAAACCAACCCAGGCGAGCATGAT 1893
 Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysLysHis 540
 Db 1894 GCAAGAAAGGACTACATCACAGCCCAAGTACATCAGAGGAGATATCGCAAGGAAGACAC 1953
 Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
 Db 1954 GCGGATAACGCGGCAAGCTTTCACAGCTTTTTCGAGGCGCTCAAAACAGAGATATTTT 2013
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 2014 GGATGTCTCAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTTGCCCAAC 2073

Db 641 TTTCTCAGTGTGTTTACAAAGAGGTTGACGACCTTTTCAAAAAACCTGATTCAGAAATATGAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATAATCTCTCTCCCTTGGACAGTTTGTGTAAGGGGACCTCAAGAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 761 GATCTGAAAAAGCCCTTTTGATAAAGCTTGGAGGACTATGAACAAAAAATAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysIleLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAGGAGAAAAAGAAACACCCCAAGCTCCATGGGATGATTCGGACCTGAATAAGCGGA 880
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTCGCGAAGAGATGGAAGAGGAGGCGCTTCTCCACCTACAGATGTGCGAG 940
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220
Db 941 TATCTGCTGAAGGTCAACGAAATCAAGATTAAGAGGAGTAGATTTACTTTCAGAAATCTG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATACTTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAGCGGTGGAAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTTCCATTGAAACGCTGTCTACGAGTCTTTCACAGATCAAGAGCCCGCAGAT 1120
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGACGCTTCATACAGCTTCGAGATATTTTGAATCCGCTTCAGGTGAA 1180
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAATTCGTTCAGAGCAGCAGCTTATAGCTTACATCAGCCTCAGGA 1240
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAAAGAACATGGAGCCGAGCGAACGCGAGCTTCAAGAAAGAGTACGCGATCCGA 1300
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle-SerHisG1 340
Db 1301 AAAGTGTGGCAGAAAAAGGAATGTTTCAGTTAAAAATGTTTTCTGACCATATTCCTATGG 1360
Qy 340 YThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPr 360
Db 1361 TACCGCTAACCCGGCTCTCTGCAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCC 1420
Qy 360 oGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaG1 380
Db 1421 TGAGGAGAAAGAGTCTTTGACCTTATTTTCATGACAGAACTTACCACTTTCAAGCTGA 1480
Qy 380 uAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLe 400
Db 1481 AGATGAACAGGATGTCAAAATATGATGCTCTGCTGCAAAATAGCAAGAAAGAGCTTT 1540
Qy 400 uAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuTh 420
Db 1541 AAACAATGCATTTAAGGGGGATGACAACTACTGGAGAAAAATAACATCGTCCAAAGAACTGAC 1600
Qy 420 rLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysG1 440
Db 1601 AAAGAGATCATCTCAGAAAGTCAGAGATGACGGGCAATGACGCTCTGTGACTGTGG 1660
Qy 440 yAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
Db 1661 GCGGCGAGATCTACATGGCTTTCACCAACTCTGGGCATCTGACCTGCATCGAGTGTTC 1720
Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVa 480

Db 1721 CGGAATCCACGAGAGCTGGGGGTTCTATTATTCAGGATGCAGTCCCTGACCTTAGATGT 1780
Qy 480 lLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluI1 500
Db 1781 ACTGGGAACATCTGAGCTGTCTGCGCAAGAAATATTGGGAATGCAGGCTTTAATAGAT 1840
Qy 500 eMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAs 520
Db 1841 CATGGAATGTTGCCCTACCAGCTGAGGACTAGTCAACCCCAACCCAGGAGCATGAA 1900
Qy 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysGlyHi 540
Db 1901 TGCAGAAGAAGACTATCATCAGCCAGTATCATCAGAGGAGATACGCAAGGAGAAGCA 1960
Qy 540 sAlaAspAsnAlaLysLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh 560
Db 1961 GCGGATAACCGCGGAGGCTTCAGTCTTTTGGAGGCCGTCAAAACGAGAGATATTTT 2020
Qy 560 eGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAs 580
Db 2021 TGGATTGCTCAAGCTTATGCTGATGTTGTTGATCTTACGGAATAAATCCCACTGCCCAA 2080
Qy 580 nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe 600
Db 2081 CGGACATGAGCGCGATGAACGGCCCTCCACTTCAGTTCAGTCCGTGGATCGAACCTC 2140
Qy 600 rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy 620
Db 2141 TCTTCACATTGTAGACTTTTATGTTTTCAGAACAGTGGGAACCTGGATTAACAGACAGGNA 2200
Qy 620 sGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLe 640
Db 2201 AGGACGACAGCCCTGCACACTCTGCTGCTGACGACAAATGCGAGTGCCTCAAGTTGCT 2260
Qy 640 uLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspI1 660
Db 2261 CCTCGGGGGAAGGCTCCATCGAGATAGAACAGAGTCAGAGAGACTCCGCTGGACAT 2320
Qy 660 eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr 680
Db 2321 TGCCNAGCCCTCAAGCAGAGACTGTGAGGAGCTGTGACCCCAAGCTTATCTGGAAG 2380
Qy 680 gPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspG1 700
Db 2381 ATTTAAATTTCTCACGTTTCAATATGAATGCGACTACTCTCCACGAAGACCTGATGA 2440
Qy 700 uSerAspAspMetAspGluLysLeuGluProSerGluAsnArgArgGluAspArgPr 720
Db 2441 AAGTGATGACGACATGGATGAGAAATTCGACGCCAGTCCCAACCGCGGGAAGACCGGCC 2500
Qy 720 oIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaAr 740
Db 2501 CATCAGCTTCTACAGCTGGGCTCCNACCAGCTTCAGTCTAAACGCTGTATCTTTGGCCAG 2560
Qy 740 gAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAs 760
Db 2561 AGATGCTGCAAACTTGGCAAGGAGAGACAGAGGCTTTTCATGCCAGCATCTTCAGAA 2620
Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSe 780
Db 2621 TGAGACTTACGAGGCCCTCTCTGAGTGGCAGGCCACCTCCCGCCAGCTGACGCCCCAG 2680
Qy 780 rThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSe 800
Db 2681 CACCACAGGCCCCCCCGCTTCTCCACGGAATGTTGGCAAGTTTCAGACAGCCCTCTC 2740
Qy 800 rAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSe 820
Db 2741 TGCTAAACACCTGTGGGAAGACAACTCTGTAAGTGTGGACGCTGGAAGCCCGCAGCGATC 2800
Qy 820 rSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAs 840
Db 2801 TTCGTGATCCGCCAGCTGTCCATCCACCGCTGCCCTCTCTTCGCGTGACATCTACCAA 2860

380 uaspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAenSerLysGluGluAlaLe 400
 1462 AGATGAACAGGGAATGCAAAATATGATGTCTGTCTGCAAAATAGCAAAAGAGAGCTTT 1521
 400 uasnAenAlaPheLysGlyAspAsnThrClyGluAenAenIleValGlnGluLeuTh 420
 1522 AAACAATGCAATTTAAGGGGATGACAATCTGGAGAAATAACATCGTCCAAAGAACTGAC 1581
 420 rlysgluIleIleSerGluValGlnArgMetThrGlyAenAepValCysAspCysG1 440
 1582 AAAGGAGATCATCTCAGAAGTGCAGAGGATGAGGGCAATGACGTCTGCTGTGACTGTGG 1641
 440 yAlaProaspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSe 460
 1642 GCGCGCAGATCCTACATGCTTCCACCAACCTGGGCATCTCTGACCTGTCATCGAGTGTTC 1701
 460 rClyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVa 480
 1702 CGGAATCCACCGAGAGCTGGGGGTTTATTTTCCAGGATGCGAGTCCCTGACCTTAGATGT 1761
 480 lleuGlyThrSerGluLeuLeuAlaLysAenIleGlyAenAlaGlyPheAsnGluI1 500
 1762 ACTGGGAACATCTGAGCTGCTCTGCCCAAGATATTTGGGAATGAGGCTTTAATGAGAT 1821
 500 eMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAs 520
 1822 CATGGAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCACCGCAGCGACATGAA 1881
 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHi 540
 1882 TGCAAGAAGGACTACATCACAGCCAGTACATCAGAGGAGATACGCAAGGAAGAAGCA 1941
 540 sAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePh 560
 1942 CGCGGATTAACGGGGGAGAGCTTACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTT 2001
 560 eGlyLeuLeuGlnAlaIatyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAs 580
 2002 TGGATTGCTCCAAAGCTTATGCTGTGATGTTGATCTTACGGAATAAAATCCCACTGCGCAA 2061
 580 nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe 600
 2062 CGGACATGAGCGGATGAACAGCGCCCTCCACCTTCAGTTCAGATCCGTGGATCGAACCTC 2121
 600 rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLy 620
 2122 TCTTCACATTTGTAGACTTTTATGTTTACAGACAGTGGGAACCTTGGATTAACAGACAGGAA 2181
 620 sGlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLe 640
 2182 AGGCAGCACAGCCCTGCACCTACTGCTGCTGACCGACAATGCGAGTGCCTCAAGTTGCT 2241
 640 uLeuArgLysAlaSerIleGluIleAlaAenGluSerGlyGluThrProLeuAspI1 660
 2242 CTTGCGGGGAGGCGCTCCATCGAGATAGCAACGAGTCCAGGAGAGACTCCGCTCGACAT 2301
 660 eAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyAr 680
 2302 TGCCGAAGCGCTTCAAGCAGAGAGACTGTGAGAGAGCTGTGACCCCAAGGCTTATCTGGAG 2361
 680 gPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspG1 700
 2362 ATTTAATCTCAGTTTACGTTGATATGATGATGATGATGATGATGATGATGATGATGATG 2421
 700 uSerAspAspMetAspGluLysLeuGlnProSerGluAenArgArgGluAspArgPr 720
 2422 AAGTGATGACGACATGATGAGAAATTCAGCCCACTCCCAACCGCGCGGGAAGACCGGCC 2481
 720 oIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaAr 740
 2482 CATCAGCTTCTACCGAGCTGGGCTCCAAACGAGCTTCAACGAGCTTCAACGAGCTTCTTTGGCCAG 2541

740 gAspAlaAlaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAs 760
 2542 AGATCTGCAAAACCTTGGCAAGGAGAGAGAGAGGCTTTTCATGCGCAGCATCTTCAGAAA 2601
 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSe 780
 2602 TGAGACTTACGAGAGCCCTCCTGAGTGGCAGCCACCTCCCGCCAGCCTTGCAGGCCCCAG 2661
 780 rThrThrSerAlaProProLeuProProArAenValGlyLysValGlnThrAlaSerSe 800
 2662 CACCACAGCGCCCGCCCGCTTCTCTCCAGGAATGTTGGCAAGATTTCAGACAGCTCTCTC 2721
 800 rAlaAenThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSe 820
 2722 TGCTAACACCTCTGGAAAGACAAACTCTGTAAAGTGGACGCTGGAAGCCGCGCAGCATC 2781
 820 rSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAs 840
 2782 TTCTGTCAGATCCGCGAGCTGTCATCCACCGCTGCCCCCTCTTCCGCTGACATCTACCA 2841
 840 nProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLe 860
 2842 TCCCTGACCCCGCAGCGCCCGCCCGCTTGGCAAGACGCCCGCGTAAATGGAAGCTT 2901
 860 uSerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPr 880
 2902 GAGCCAGCGAGCAGAGCTGCGCCCGCTGGGATCTTCACAGATCAGCGCCCGCCACCTCTGCC 2961
 880 oProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerTh 900
 2962 CCCAGCGCCCGCAGCGCCCTCCCGCAAGAGAGCTTCCGCGGGGGCTGACAAGTCCAC 3021
 900 rProLeuThrAsnLysGlyGlnProArGlyProValAspLeuSerAlaThrGluAlaLe 920
 3022 CCCACTGACCAACAAAGCCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAAGCTCT 3081
 920 uGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940
 3082 GGGTCTCTGTCCAATGCTATGCTGCTGCGAGCCCGCTGCAACCATGCTGTGGCTGACAACCC 3141
 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAenCysValAlaAspAsnPr 960
 3142 GGCAACCAAGTTGAAGCCTGAGCGGGTGAAGAGCGCTCTATACTGTGTGGCTGACAACCC 3201
 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980
 3202 CGATGAGCTCACCCTTCTCCGAGGGGATGTGATCATCTGTCGCGGGAGGAGGACAGGA 3261
 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
 3262 GTGGTGGATTGGCCACATTTGATGGAGATCTCTGTCGCAAGCGCATTTCCCGGTGTCAAT 3321
 1000 eValHisPheIleAlaAsp 1006
 3322 TGTGACATTTATCGCTGAC 3340

RESULT 8
 ; Sequence 6642, Application US/60278232
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, MacDonald
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Diep, Dinh
 ; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
 ; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
 ; FILE REFERENCE: GX-0011 P
 ; CURRENT APPLICATION NUMBER: US/60/278,232
 ; CURRENT FILING DATE: 2001-03-30
 ; NUMBER OF SEQ ID NOS: 12,557
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 6642
 ; LENGTH: 5810

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 331279.4
US-60-278-232-6642

Alignment Scores:
Pred. No.: 1,869-216 Length: 5810
Score: 5236.00 Matches: 1002
Percent Similarity: 99.60% Conservative: 1
Best Local Similarity: 99.50% Mismatches: 3
Query Match: 99.30% Indels: 1
DB: 97 Gaps: 0

US-09-914-042-1 (1-1006) x US-60-278-232-6642 (1-5810)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleAla 20
Db 322 ATGCGGACACAGATCTCGTGTGCGGAATTCTGTGGCCGAGACCCATGAGGACTACAGGCG 381
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 382 CCCACGGCTCCAGCTTCACCACCCGCGCAGCGGCGAGTCCCGGAAACACTGTGCGCGCCATC 441
Qy 41 GluGluAlaLeuAsp-ValAspAtqMetValIleuTyrIleValMetIleValIleValAl 60
Db 442 GAGGAGGCTTTGGACCGTGGACCGGATGGTCTTTTACAAATGAAGAAATCCGTGAAGC 501
Qy 60 aIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuG1 80
Db 502 AATCAACAGCTCTGGGCTGGCTCAGTGTGGAATGAAGACGAGTACACCCAGGCTCTGA 561
Qy 80 uIlePheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLy 100
Db 562 GAAGTTTGGCGCAACTGTGTATGAGAGATGACCCAGATTTAGGAAGTGGTGTCTCTCAA 621
Qy 100 sPheSerValPheThrIleValLeuThrAlaLeuPheIleValAsnLeuIleGlnAsnMetAs 120
Db 622 GTTCTCAGTGTTTACAAAGAGATTGACAGCACTTTTCAAAACCTTGATTCAGAAATATGAA 681
Qy 120 nAsnIleIleSerPheProLeuAspSerLeuLeuIleGlyAspLeuIleGlyValIleG1 140
Db 682 CAACATAATCTCTTCCCTTTGGACAGTTTGTGAAGGGGACCTGGAAGGAGTGAAAGG 741
Qy 140 yAspLeuIleValProPheAspLysAlaTrrIleAspTyrGluThrIleIleThrLysI1 160
Db 742 GGATCTGAAAAAGCCCTTTTGATAAAGCTTGAAGGAGCTATGAAAACAAAAATAACCAAGAT 801
Qy 160 eGluIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 180
Db 802 AGAAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 861
Qy 180 yAlaGluIleAlaGluGluMetGluIleGluArgArgPhePheGlnLeuGlnMetCysG1 200
Db 862 AGCGGAAATTGCCGAAGAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 921
Qy 200 uTyrIleuLeuIleValAsnGluIleIleIleIleIleIleIleIleIleIleIleIleIle 220
Db 922 GTATCTGCTGAAGGTCAACGAAATCAAGATTAAGAAAGGAGGTAGATTTTACTTCAGAAATCT 981
Qy 220 uIleIleValPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuIleValAlaValIle 240
Db 982 GATCAATACCTTTTATGCCCCTAATGCAATTTTTTTCAGGATGGACTCAAGCGCGTGGAAAG 1041
Qy 240 rLeuIleProSerIleGluThrLeuSerThrAspLeuHisThrIleIleGlnAlaGlnAs 260
Db 1042 CCTCAACCTTCCATTGAACGCTGTCTACGGATCTTCACAGATCAACAGGCGCCAGGA 1101
Qy 260 pGluGluArgGlnLeuIleGlnLeuArgAspIleIleuIleuIleuIleuIleuIleuIle 280
Db 1102 TGAAGAAAGAGGCGAGTTGATACAGCTTCGAGATATTTTGAAGATCCGCAATTCAGGTTGA 1161
```

```
Qy 280 uGlnIleValAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnG1 300
Db 1162 ACAGAAAGAGGACTCCCAATTCGTAGAGACACAGCTTATAGCTTACATCAGCTCAGG 1221
Qy 300 yAsnIleGluHisGlyThrGluArgAsnGlySerLeuTyrIleIleValIleAsr 320
Db 1222 AAACAAGGAAACATGGGACCGGACCGGAGCTCTTACAAAGAGAGTACGGGATCCG 1281
Qy 320 gIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 340
Db 1282 AAAAGTGTGGCAGAAAAGGAAATGTTTCACTTAAATAATGGTTTCTGACCATATCCCATGG 1341
Qy 340 yThrAlaAsnArgProProAlaIleValLeuAsnLeuLeuThrCysGlnValIleThrAsnPr 360
Db 1342 TACCGCTAACCGGCTCTCTGCAAGCTCACTGCTAACCTGCCAGGTGAAGACCAACCC 1401
Qy 360 oGluIleValIleValIleValIleValIleValIleValIleValIleValIleValIle 380
Db 1402 TGAGGAGAGAGAGTGTCTTACCTTATTTACATGACAGAACTTACCACTTTCAAGCTGA 1461
Qy 380 uAspGluGlnCysGlnIleTrrMetSerValIleGlnAsnSerIleValIleAlaLe 400
Db 1462 AGATGAACAGGAATGTCAATATATGATGTCTGTCTGCAAAATAGCAAAAGAAAGCTTT 1521
Qy 400 uAsnAsnAlaPheIleValIleValIleValIleValIleValIleValIleValIle 420
Db 1522 AAACAATGCAATTAAGGGGATGACATATCTGGAGAAAATAACATCTGCCAAGAACTGAC 1581
Qy 420 rLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysG1 440
Db 1582 AAAGGAGATCATCTCAGAAAGTGCAGAGATGACGGCAATGACGCTCTGTGACTGTGG 1641
Qy 440 yAlaProAspProThrTrpIleuSerThrAsnLeuGlyIleLeuThrCysIleGluCysE 460
Db 1642 GGCCCAAGATCTACATGGCTTCCCAACCTGGGCATCTCGACCTGCACTGAGTGTTC 1701
Qy 460 rGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVa 480
Db 1702 CGGAATCCACGAGAGCTGGGGTTCATATTCCAGGATGCACTCCCTGACCTTAGATGT 1761
Qy 480 lLeuGlyThrSerGluLeuLeuAlaIleAsnIleGlyAsnAlaGlyPheAsnGluI1 500
Db 1762 ACTGGAAACATCTCAGCTGTCTCGCAAGATATTGGGAATTCAGGCTTTAATAGAT 1821
Qy 500 eMetGluCysCysLeuProAlaGluAspSerValIleProAsnProGlySerAspMetAs 520
Db 1822 CATGGAATGTTCCTTACAGCTGAGGACTCAGTCAAAACCCCAACCCAGGACGAGATGAA 1881
Qy 520 nAlaArgLysAspTyrIleThrAlaIleValIleGluArgArgTyrAlaArgLysIleHi 540
Db 1882 TGCAAGAAAGGACTACATCACAGCCAAAGTACATCGAGAGAGAGATACGCAAGGAAGAGCA 1941
Qy 540 sAlaAspAsnAlaAlaIleValIleValIleValIleValIleValIleValIleValIle 560
Db 1942 CGCGGATAACCGCGGAGCTTCACAGTCTTTGGAGGCGCTCAAAACGAGAGATATTTT 2001
Qy 560 eGlyLeuLeuGlnAlaTyrAlaAspGlyValIleAspLeuThrGluIleValIleProLeuAlaAs 580
Db 2002 TGGATGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGGCCAA 2061
Qy 580 nGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSe 600
Db 2062 CGGACATGAGCGGATGAACGGCCCTCCACCTTTCAGTTCAGATCCGTGGATCGAACCTC 2121
Qy 600 rLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspIleThrGlyLy 620
Db 2122 TCTTTCATATTGACATTTTATGTTTCAGAAACAGTGGGAAACCTGGATTAACAGAGAGGAA 2181
Qy 620 sGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuIleValLe 640
Db 2182 AGGAGAGACAGCCCTGCACTACTCTGCTGACGACAAATCCGAGTGCCTCAAGTTGCT 2241
Qy 640 uLeuArgGlyIleValAsnIleGluIleAlaAsnGluSerGlyGluThrProLeuAspI1 660
```



```
Db 2242 CTGCGGGGGAAGGCTTCATCGAGATAGCAACGAGTCAGGAGAGACTCCGCTGGACAT 2301
Qy 660 ealalyArGLeuLysHisGluHisCysGluGluLeuThrGlnAlaLeuSerGlyAr 680
Db 2302 TGCCAAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGCTGTGACCCAGCCCTTATCTGGAAG 2361
Qy 680 gheAenSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuSpG1 700
Db 2362 ATTAAATCTCAGCTTCAGCTTGAATATGAATGGGAGCTACTCCAGAGAGCTGGATGA 2421
Qy 700 uSerAspAspMetAspGluLysLeuGlnProSerGluAenArgArgGluAspArgPr 720
Db 2422 AAGTATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2481
Qy 720 oileSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaAr 740
Db 2482 CATCAGCTTACCAGCTGGGCTCCAAACAGCTTCAGTCTAAGCTGTATCTTTGGCCAG 2541
Qy 740 gAspAlaAlaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAs 760
Db 2542 AGATGCTGCAACCTTGCAGAGAGAAGCAGAGGGCTTTCATGCCAGCATCTTGCAAA 2601
Qy 760 nGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSe 780
Db 2602 TGAGACTTACGGAGGCTCTCCTGAGTGGCAGCCACCTCCCGCCAGCTGCGAGCCCCAG 2661
Qy 780 rThrThrSerAlaProProLeuProProArgAenValGlyLysValGlnThrAlaSerSe 800
Db 2662 CACCACAGGCGCCCGCTTCTCCAGGATGTTGCAAGATTTCAGACAGCTCTCTC 2721
Qy 800 rAlaAenThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSe 820
Db 2722 TGCTAACACCTGTGGAGACAACTCTGTAAGTGTGGACGTGGAAGCGCGCAGCATC 2781
Qy 820 rSerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAs 840
Db 2782 TTCGTGATGCGCCAGCTGTCCATCCCGCTGCGCCCTCTTCGCGTGACATCTACCAC 2841
Qy 840 nProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLe 860
Db 2842 TCCCTGTACCCCGCCAGCGCCCGCTTCCAGGATCTCCAGAGCGCCCGGTTAAGAGCCTT 2901
Qy 860 uSerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPr 880
Db 2902 GAGCAGCGCAGCAAGCCTGCGCCGCTGGGATCTCACAGATCAGCGCCCGCCACCTTGCC 2961
Qy 880 oProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerTh 900
Db 2962 CCCACAGCGCGCCAGCGCGCTCCCGCAGAAGAAGCTGCGCGGGGGCTGACAAGTCCAC 3021
Qy 900 rProLeuThrAenLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLe 920
Db 3022 CCCACTGACCAACAAAGGCCAACCGAGAGACCTGTGGATCTCTCTGCAACGGAAGCTCT 3081
Qy 920 uGlyProLeuSerAenAlaMetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 3082 GGGTCTCTGTCCATGTATGTCTCTGAGCGCCCTGACCCATGCTCTAGGAAGTGCAC 3141
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAenCysValAlaAspAsnPr 960
Db 3142 GGCACCAAGTTGAAGCCTGAGCGGGTGAAGCGCTCTATTAACCTGTGTGGCTGACAACCC 3201
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValLeuAspGlyGluGluAspGlnCl 980
Db 3202 CGATGAGCTCACCTTCTCCAGGGGATGTGTATCATCTGTGGACGGGGAGGAGACAGGA 3261
Qy 980 uTyrTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 3262 GTGGTGGATTGGCCACATTTGATGAGATCCTGTGTCGAAGGCGCATTTCCCGGTCTATT 3321
Qy 1000 eValHisPheIleAlaAsp 1006
|||||
```

```
Db 3322 TGTGCATTTTATCGCTGAC 3340
RESULT 9
US-09-572-411-6530
; Sequence 6530, Application US/09572411
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Xu, Yong Yao
; APPLICANT: Acton, Susan L.
; APPLICANT: Shyjan, Andrew W.
; APPLICANT: Fraser, Christopher C.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USBS
; FILE REFERENCE: 1600.1091-001
; CURRENT APPLICATION NUMBER: US/09/572,411
; EARLIER FILING DATE: 2000-05-12
; EARLIER APPLICATION NUMBER: 60/133,993
; EARLIER FILING DATE: 1999-05-13
; EARLIER APPLICATION NUMBER: 60/135,617
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,816
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,817
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,950
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,679
; EARLIER FILING DATE: 1999-05-24
; EARLIER APPLICATION NUMBER: 60/135,630
; EARLIER FILING DATE: 1999-05-24
; NUMBER OF SEQ ID NOS: 7417
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6530
; LENGTH: 6389
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(6389)
; OTHER INFORMATION: n = A,T,C or G
US-09-572-411-6530

Alignment Scores:
Pred. No.: 2,7e-215 Length: 6389
Score: 5210.00 Matches: 996
Percent Similarity: 99.11% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 9
Query Match: 98.81% Indels: 0
DB: 26 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-572-411-6530 (1-6389)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATCGCGGACCAAGATCTCCGCTGCGGAATTCGTGGCCGAGACCCATGAGACTACAGGGC 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGGCTCCAGCTTACCACCGCCGCGAGTCCGCGGACCACTGTGGCGGCATC 460
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 461 GAGGAGGGTTTGGAGCTGGACCGGATGGTTCTTTACAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAenSerSerGlyLeuAlaHisValGluAenGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGGCTGGCTCACGTGAAAATGAAGAGCAGTACACCCAGCTCTGGAG 580
Qy 81 LysPheGlyGlyAenCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGACAGATGACCCAGATTTAGGAAGTGCCTTCTTGAG 640
```

101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db
641 TTCTCAGTGTGTTTCAAGAGAGTTGACAGCACTTTTCAAAAAACCTGATTCAGAAATATGAAC 700
Qy
121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db
701 AACATAATCTCTTCCCTTTGGACAGTTTGTGAAGGGGGACCTGAAAGGAGTGAAGGG 760
Qy
141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db
761 GATCTGAAAAAGCCCTTTTGATTAAGCTTTGGAAGGACTATGAACAAAAAATAACCAAGATA 820
Qy
161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysIleSerGly 180
Db
821 GAAAGGGAAGAAAGGAACACGCCNAGCTCCATGGGATGATTGGGACTGAATAAGCCGA 880
Qy
181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db
881 GCGGAATTTGCCGAAGAGATGGAAGAGGAGAGGGCGCTTCTCCAGCTACAGATGTGCGAG 940
Qy
201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db
941 TATCTGCTGAAGGTCAACGAAATCAAGATTAAAAAGGGAGTAGATTTTACTTCAGAAATCTG 1000
Qy
221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db
1001 ATCAAAATACTTTTCATGCCCAATGCATTTTTTTTTCAGGATGGACTCAAGCCGTGGAAGC 1060
Qy
241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db
1061 CTCAAACTTCCATTGAAACGCTGTCTACGGATCTTCACAGATCAAAACAGGCCCCAGGAT 1120
Qy
261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db
1121 GAAGAAAGAGCGCAGTTGATACAGCTTCGAGATATTTTGAATCCGCAATTCAGGTTGAA 1180
Qy
281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db
1181 CAGAAAGAGGACTCCCAATTCGTGAGAGCAGCAGCTTATAGCTTACATCAGCCTCAGGGA 1240
Qy
301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db
1241 AACAAAGAAATGCGACCGAGCGGAACGGCAGCCTCTCAAGAGAGTGACGGGATCGA 1300
Qy
321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db
1301 AAAGTGTGGCAGAAAGGAAATGTTTCAGTTAAAAATGGTTTTCTGACCATATCCCATGGT 1360
Qy
341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db
1361 ACCGCTAACCGGCCCTCTGCAAGCTCAACCTGCTTAACCTGCCAGGTGAAACCAACCTT 1420
Qy
361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db
1421 GAGCAGAAAGAGTGTCTTTCAGCTTATTTTCAATGACAGAACTTACCACTTTCAAGCTGAA 1480
Qy
381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db
1481 GATGAACAGGAATGTCAAAATATGGATGTCTGTGCTGCAAAAATAGCAAAAGAAAGCTTTA 1540
Qy
401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db
1541 AACAAATGCAATTAAGGGGGATGACAAATCTGGAGAAATTAATCATGCTCCAGAACTGACA 1600
Qy
421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db
1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGGCAATGACGTCTGTGACTGTGGG 1660
Qy
441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db
1661 GCGCAGATCCTACATGCGCTTTCCCAACCTGGGCATCTCTGACCTGCATCGAGTGTTC 1720
Qy
461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480

1721 GGAATCCACCGAGAGCTGGGGTTCATTATTTCCAGGATGCAGTCCCTGACCTTAGATGTA 1780
Qy
481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db
1781 CTGGGAACATCTGAGCTGCTGCTCCCAAGAAATATTTGGGAATGCAAGGCTTTTAATGAGATC 1840
Qy
501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db
1841 ATGGAATGTTCCTTACCAGCTGAGACTCAGTCAAAACCCAAACCCAGGCGAGCATGAAAT 1900
Qy
521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db
1901 GCAAGAAAGGACTACATCACAGCCAGTACATCCAGAGGAGATACGCAAGGAAGAGCAC 1960
Qy
541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db
1961 GCGGATAACGCGCGAAGCTTTCACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTTTTT 2020
Qy
561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db
2021 GGATGTCTCCAAAGCTTATGCTGATGGTGGATCTTACGGAATAAATCCCCACTGGCCAAAC 2080
Qy
581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db
2081 GGACATGCGCGGATGAAACGCGCTCCACCTTCGAGTCAGATCCGTGGATCGAACCTCT 2140
Qy
601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db
2141 CTTTCACATTTAGACTTTTTCAGAACAGTGGAAACCTGGATAAAACAGACAGGGAAA 2200
Qy
621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db
2201 GGCAGCAGCGCTTGCACACTCTGCTGCTGACCCACAATGCCGAGTGCCTCAAGTTGCTC 2260
Qy
641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db
2261 CTGCGGGGAAGGCTCCATCGAGATAGCAAAACAGTCAGGAGAGACTCCCTCGACATT 2320
Qy
661 AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db
2321 GCCAAGCGCTCAAGCACGAGCACTGTGAGGAGCTGTGACCAAGCCTTATCTGGAAGA 2380
Qy
681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db
2381 TTTAAATCTCACGTTCAGTTGAATATGAATGGCGACTACTCCACGAAGAACTCGATGAA 2440
Qy
701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db
2441 AGTGATGACGACATGGATGAGAAATTCAGCGCCAGTCCCAACCGCGGGAAGACCGGCC 2500
Qy
721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db
2501 ATCAGCTTCTACCAAGCTGGGCTCCAAACAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
Qy
741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db
2561 GATGCTGCAAAACCTTGCAGGAGAGCAGAGGCTTTTCATGCCAGCATCTTTCAGAAAT 2620
Qy
761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db
2621 GAGACTTACGGAGCGCTCCTGAGTGGCAGCCCACTCCCGCCAGCTGCAGCCCCCAGC 2680
Qy
781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db
2681 ACCACCGCGCCCCCGCTTCTCCACGGAATGTTGGCAAAAGTTTCAGACAGCTCTCTCT 2740
Qy
801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db
2741 GCTAACACCTCTGTGGAGACAAACTCTGTAAAGTGGCTGCTGCTAGCGCGCGCGCGCCCC 2800
Qy
821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
::: |||||


```
Db 2801 CGGTGAGATCGCCAGCTGTGATCCACCGCTGCCCTCTTCGGGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCACACGCCGCCACCCGTTGCCAAGACGCCAGGTAATGGAGCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnLysArgProProProLeuPro 880
Db 2921 AGCCAGCGAGCAAGCCGCGCTGCGGCTGCGATCTCACAGATCAGGCCGCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACGCGCGCCAGCGGCTCCGCGAAGAGCCCTGCGCGGGGCTGACAGTCCACC 3040
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTCTCTGCAACGGAGCTCTG 3100
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCAATGCTATGGTCTGCGAGGCCCTGCAACCCCTGCAACGCTGCGAG 3160
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuThrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAACCCC 3220
Qy 961 AspGluLeuThrPheSerGluGlyAspValLysLysValLysLysValLysLysGlu 980
Db 3221 GATGAGCTCACCTCTCCAGGGGATGTGATCATCTGTGGAGCGGGAGGAGGACGAG 3280
Qy 981 TrpThrLysGlyHisLysAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTGGATGGCCACATTGATGGAGATCTGTGTGCGAAAGCGCATTCCTCCGCTGTCA 3340
Qy 1001 ValHisPheLysAlaAsp 1006
Db 3341 GTGCACITTTATCGCTGAC 3358

RESULT 10
US-10-170-235-8470
; Sequence 8470, Application US/10170235
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; FILE OF INVENTION: TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF
; FILE REFERENCE: CL001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 8470
; LENGTH: 5670
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-8470

Alignment Scores:
Pred. No.: 7,78e-213 Length: 5670
Score: 5152.00 Matches: 988
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 2
Query Match: 97.71% Indels: 16
DB: 49 Gaps: 1

US-09-914-042-1 (1-1006) x US-10-170-235-8470 (1-5670)

Qy 1 MetProAspGlnLysSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCGGACCAAGATCTCGGTGCGAATTCGTGGCCGAGACCCATGAGGACTACAAAGCG 400
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaLys 40
Db 401 CCCAGCGCTCCAGCTTCACCAACCGGCGGCGGAGTCCCGGAACACTGTGGCGGCCATC 460
```

```
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 461 GAGGAGGCTTTGGACCTGCGACCGGATGTTCTTTACAAAATGAAGAAATCCGTGAAGCA 520
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCAGCTGGAAATGAAGAGCAGTACACCAGGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTGGCTCTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysLysLysMetAsn 120
Db 641 TTCTCAGTGTTTACAAAGGAGCTTGACAGCACTTTTCAAAAACCTGATTCAGATATGAAC 700
Qy 121 AsnLysLysSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTCTGAAGGGGGACCTGAAAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaThrLysAspTyrGluThrLysLysLysLysLys 160
Db 761 GATCTGAAAAAGCCTTTTGATAAAGCTTGGAAAGGACTATGAAACAAAATAAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetLysLysLysLysLysLys 180
Db 821 GAAAGGAGAAAAAGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATAAGCGGA 880
Qy 181 AlaGluLysLysGluLysGluMetGluLysGluLysGluLysGluLysGluLysMetCysGlu 200
Db 881 GCGGAAATTTGCGAAGAGATGGAAAGGAGAGGCGCTCTTCCAGCTACAGATGTGGCAG 940
Qy 201 TyrLeuLysLysValAsnGluLysLysLysLysLysLysLysLysLysLysLysLys 220
Db 941 TATCTGCTGAAGTCAACGAATCAAGATTTAAAAGGAGTAGATTTACTTTCAGATCTG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAATACTTTTCATGCCAATGCAATTTTTTTCAGGATGGACTCAAAAGCGTGAAGAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrLysLysGluAlaGlnAsp 260
Db 1061 CTCAAACCTTCCATTTGAAACCGCTGTACAGGATCTTCACAGATCAAAACAGGCCAGGAT 1120
Qy 261 GluGluArgArgGlnLeuLysGlnLeuArgAspLysLysLysLysLysLysLysLysLys 280
Db 1121 GAAGAAGAGGAGGAGTGTATACAGCTTCGAGATTTTGAATTCGCAATTCGAGGTTGAA 1180
Qy 281 GlnLysGluAspSerGlnLysArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTGAGAGCACAGCTTATAGCTTACATCAGCCTCAGGGA 1240
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLysArg 320
Db 1241 AACAAAGGAACATGGGACCGAGCGGACCGGACCGCTCTACAAAGAGAGTGAAGGAGTCCGA 1300
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrLysSerHisGly 340
Db 1301 AAAGTGTGGCAGAAAGGAATGTTCAAGTTAAAATGGTTTTCTGACCATATCCCATGGT 1360
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGCTAACCTGCCAGGTGAAGACCAACCT 1420
Qy 361 GluGluLysLysCysPheAspLeuLysSerHisAspArgThrThrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAGAGAGTCTTTGACCTTATTTTACATGACAGAACTTACCACCTTTCAAGCTGAA 1480
Qy 381 AspGluGlnGluCysGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 400
Db 1481 GATGAACAGGAATGTCAATATGATGTCTGTCTGCAAAATGCAAAAGAGAGAGCTTTA 1540
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
```

```
Db 1541 AACATGCTTTAAGGGGATGACATCTGGAGAAATAACATCGTCCAAAGACTGACA 1600
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGGAGATCATCTCAGAAAGTCAGAGGATGACGGCAATGACGCTCTGCTGTGACTGTGG 1660
Qy 441 AlaProAspProThrTyrLeuSerThrAsnLeuGluIleLeuThrCysIleGluCysSer 460
Db 1661 CGCGCAGATCCTACATGCTTTTCCACCAACCTGGGCATCCTGACCTGCATCGAGTGTTC 1720
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1721 GGAATCCACGAGAGCTGGGGTTCATTTATCCAGGATGCACTCCCTGACCTTAGATGA 1780
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGGACATCTGAGCTGCTGCTGCCAAGAAATATTGGGAATGCGAGGCTTTAATGAGATC 1840
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATGTTCCCTACCAAGCTGAGGACTCAGTCAAAACCCCAAGGAGGACATGAAT 1900
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAAGGACTCATCATCAGCCCAAGTACATCGAGAGGATACGCAAGGAAGACAC 1960
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 CGCGATAACGCGCGAAGCTTCACAGTCTTTGCGAGGCGGTCAAAACGAGAGATATTTT 2020
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATGTGCTCCAAAGCTTATGCTGATGGTGTGATCTTTACGGAAAAAATCCCACTGGCC 2080
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACCTTGCACTGAGTCAGATCCGTGGATCGAACCTCT 2140
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2141 CTTCACTGTGATGCTTTTGTAGTTTCTCAGAACAGTGGGAACTTGATAAACACAGACAGGAAA 2200
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2201 GGCAGCAGACGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2260
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2261 CTGCGGGGAGGCGCTCCATCGAGATAGCAACGAGTCAAGAGACACTCCGCTGGACATTT 2320
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCTCAACGACGAGCAGCTGTGAGGAGCTGTGACCCAAAGCTTTATCTGGAAGA 2380
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTAATTTCACTGCTTCACTGCTTCACTGCTTCACTGCTTCACTGCTTCACTGCTTCACTGCT 2440
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2441 AGTGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCACTGCTTCACTGCTTCACTGCTTCACTGCT 2560
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAACCTTGGCAAGGAGAGCAGAGGCTTTTCATGCCAGGATCTTTGAGAAAT 2620
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
```

```
Db 2621 GAGACTTACGAGGCGCTCTCTGAGTGGCAGCCCACTCCCGCCAGCTGCAGCCCCCAGC 2680
Qy ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACGAGCGCCCCCGCTTCTCTCCAGCGAATGTTGGCAAAGTTCAGACAGCTCTCTCT 2740
Qy 801 AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTCTGTGGAGACAACTCTGTAAAGTGTGGACGTTGAAGCCGAGCGGATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCGCAGCTGTCCATCCAGCGCTGCCCTCTCTTCGCGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGCAGCGCCCCCACCCTGTTGCCAAGACGCCCGCAGCGTAATGGAAGCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGGAGCAAGCTTGGCCCGCTGGGATCTCACAGATCAGCGCCCCCACCTCTGCCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCGCAGCGCTCCCGCAGAGAAGCTTGGCCGCGGA----- 3025
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3026 -----GTGGATCTCTCTGCAACGGAAGCTCTG 3052
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3053 GGTCTCTCTGTCATGCTATGCTCTGCGAGCCCCCTGCACCATGCTCTAGGAAGTCGCGAG 3112
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3113 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATAACTGTGTGCTGACAACCCC 3172
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGlu 980
Db 3173 GATGAGCTCACTTCTCCGAGGGGATGTGATCATCTGTTGGAGCGGGAGGAGGACGAGGAG 3232
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyValAlaPheProValSerPhe 1000
Db 3233 TGGTGGATTGGCCACATGATGGAGATCTCTGTCGAAAGCGCATTTCCCGGTGTCTATT 3292
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3293 GTGCACITTTATCGCTGAC 3310
```

RESULT 11

```
US-60-452-680-5751
; Sequence 5751, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5751
; LENGTH: 5670
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-5751

Alignment Scores: 1.41e-212 Length: 5670
Pred. No.: 5146.00 Matches: 987
Score: 98.11%
Percent Similarity: 98.11% Conservative: 0
```

Best Local Similarity:	98.11%	Mismatches:	3
Query Match:	97.59%	Indels:	16
DB:	115	Gaps:	1
US-09-914-042-1 (1-1006) x US-60-452-680-5751 (1-5670)			
Qy	1	MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla	20
Db	341	ATCCGCGACCAAGATCTCCGTGTCGGAATTCGTGGCCGAGACCATTAGAGACTACAAAGCG	400
Qy	21	ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle	40
Db	401	CCACGGCTCCAGCTTCCACCCCGCAGCGCCAGTCCCGGAAACACATGTGGCGCCATC	460
Qy	41	GluGluAlaLeuAspValAspArgMetValLeuTyrIysMetLysLysSerValIysAla	60
Db	461	GAGGAGGCTTTGGAYGTGGACCGGATGTTCTTTACAAAATGAAGAAATCCGTGAAGCA	520
Qy	61	IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu	80
Db	521	ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGAGCAGTACACCCAGGCTCTGGAG	580
Qy	81	LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys	100
Db	581	AAGTTTGGCGGCACTGTGTATGCAGAGATGCCAGATTTTAGGAAGTGCCTCTGAAG	640
Qy	101	PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn	120
Db	641	TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCANAAACCTGATTCAGATATGAAC	700
Qy	121	AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly	140
Db	701	AACATAATCTCCTTCCCTTTGGACAGTTTGTCTGAAGGGGGAGCTCAAAAGGAGTGAAGGG	760
Qy	141	AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle	160
Db	761	GATCTGAAGAAAGCCCTTTTGATAAAGCTTGGAGGAGCTATGAACAAAATAAACCAAGATA	820
Qy	161	GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly	180
Db	821	GAAAGGAGAGAAAAGGACACGCCAAGCTCCATGGGATGATTCGGACTGAAATTAAGCGGA	880
Qy	181	AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu	200
Db	881	CGCGAAATTCGCGAAGAGATGAAAGAGAGAGCGCTCTTCCAGCTACAGATGTGCGAG	940
Qy	201	TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu	220
Db	941	TATCTGCTGAAGGTCACGAAATCAAGATTAAGAGGGAGTAGATTTTACTTCAGAAATCTG	1000
Qy	221	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
Db	1001	ATCAAAATCTTCATGCCCAATGCAATTTTTCAGGATGACTCAAGCCGTGGAAAGC	1060
Qy	241	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
Db	1061	CTCAAACTCTCCATTGAAACGCTGTACGGATCTTCACAGATCAAAACAGGCCCGGAT	1120
Qy	261	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
Db	1121	GAAGAAAGAGGAGCTGATACAGCTTCGAGATATTTTGAATCCGCATTGCAAGTTGAA	1180
Qy	281	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
Db	1181	CAGAAAGAGGACTCCCAATTCGTACAGACACAGCTTATAGCTTACATCAGCCTCAGGGA	1240
Qy	301	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
Db	1241	AACAAGGAACATGGGACCGAGCGGAACCGCAGCTCTTCAAGAAAGAGTGCAGGATCCGA	1300
Qy	321	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
Db	1301	AAAGTGTGGCAGAAAGGAAATGTTTCAGTTAAATAATGTTTCTGACCATATATCCATGCT	1360

Qy	341	ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
Db	1361	ACCGCTAACCGCGCTCTCTGCAAGCTCAACTGCTAACTGCGAGTGAAGACCAACCT	1420
Qy	361	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
Db	1421	GAGGAGAAAGAGTGTCTTTGACCTTATTTTCACTGACAGAACTTACCACCTTTCAAGCTGAA	1480
Qy	381	AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400
Db	1481	GATGACAGGAATGTCAANTATGATGCTGTCTGTCGAAATAGCAAGAGAGACCTTTA	1540
Qy	401	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
Db	1541	AACAATGCTTAAAGGGGATGACAACTACTGGAGAAATAACATCGTCCAGAACTGACA	1600
Qy	421	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
Db	1601	AAGGAGATCATCTCAGAAAGTGCAGAGGATGACGGCAATGAYGTCTGTGTGACTGTGGG	1660
Qy	441	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
Db	1661	CGCCAGATCTCTACATGGCTTTCCACCAACTCTGGGATCTTGACCTGCATCGAGTGTCC	1720
Qy	461	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
Db	1721	GGATCCACCCAGAGCTGGGGGTTTCAATTTATCCAGGATGCAGTCTCGACCTTAGATGTA	1780
Qy	481	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
Db	1781	CTGGGAACATCTGAGCTGCTCGCCAAAGAAATATTTGGAAATGCAAGGCTTAATGAGATC	1840
Qy	501	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
Db	1841	ATGGATTTGCTTACCCAGCTGAGACTCAGTCAAAACCCAGGAGGAGGACATGAT	1900
Qy	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis	540
Db	1901	GCAAGAAAGGACTACATCACAGCCCAAGTATATTCAGAGAGAGATACGCAAGGAAGACAC	1960
Qy	541	AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe	560
Db	1961	CGCGATAACGCGGCGAGCTTTCACAGCTTTTTCGAGGCGCTCAAAACAGAGATATTTT	2020
Qy	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn	580
Db	2021	GGATTTCTCCAAAGCTTATGCTGATGGTGGATCTTACGGAATAAATCCCTGCTGCCAAC	2080
Qy	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
Db	2081	GGACATGAGCCGGATGAAACGGCCCTCCACCTTTCGCTGATCGTGGATCGAACCCTCT	2140
Qy	601	LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys	620
Db	2141	CTTCACATTTGATGACTTTTATAGTTTCAGAACAGTGGGAACTTGGATAAACAAGAGGAAA	2200
Qy	621	GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu	640
Db	2201	GGCAGCAGCCCTGCACTACTGCTGCTGCCCAATGCCGAGTGGCTCAAGTTGCTC	2260
Qy	641	LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle	660
Db	2261	CTCGGGGAGGAGCTCCATCGATGACAAAYAGTGCAGGAGAGACTCCCTGACATT	2320
Qy	661	AlaLysArgLeuLysHisGluHisCysGluLeuLeuThrGlnAlaLeuSerGlyArg	680
Db	2321	GCCAAAGGCGCTCAAGACAGCAGCACTGTGAGGAGCTGTGACCAAGCCTTATCTGGAAGA	2380
Qy	681	PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu	700
Db	2381	TTTAATTTCTACGTTTCAGTTTGAATATGAAATGGCGACTACTCCACGAAAGACCTGGATGAA	2440

```
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
Db 2441 AGTGATGACGACATGGATGAGAAATTGCAGCCAGTCCCAACCGCGGGAAGACGGGCC 2500
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACCAAGTGGCTCCCAACCAAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAAACTTTGCCAAGASAAAGCAGAGGGCTTTTCATGCCAGCATCTTGCAGAA 2620
Qy 761 GluThrTyrGlyAlaLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2621 GAGACTTACGAGAGCCCTCTGAGTGGAGCCACCCTCCCGCCAGCCCTGCAGCCCCAGC 2680
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACGAGCGCCCGCCGCTTCTCCACGGAATGTTGGCAAAAGTTCAGACAGCCCTCCTCT 2740
Qy 801 AlaAsnThrLeuTyrPheThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTGTGGAAGACAACTCTGTAAGTGTGGACGGTGGGAAGCGCGCAGCGATCT 2800
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCACAGCTGTCAATCCACGCTGCCCTCTTTCGGGTGACATCTACCAAT 2860
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCCACGCGCCGCCACCCCTGTCCAAAGACGCCAGCGTAATGGAGCCCTTG 2920
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
Db 2921 AGCCAGCGCAGCAAGCCCTGCCCGCTGGGATCTCACAGATCAGGCCGCCACCTCTGCC 2980
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCCCGCCAGCGCGCTCCCGCAGAAAGACCTGCCCGCGGA----- 3025
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3026 -----GTGGATCTCTCTGCAACGGAAGCTCTG 3052
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3053 GGTCCTCTGTCCAATGTCTATGCTCTGCAGCCGCCCTGCACCCCATGCTAGGAAGTCGCAG 3112
Qy 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3113 GCAACCAAGTTGAAGCCTAAGCGGTGAAGCGCTCTATACTGTGTGGCTGACAAACCCC 3172
Qy 961 AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu 980
Db 3173 GATGAGCTCACCTTCTCCAGGGGATGTGATCATCGTGGAGCGGGAGGAGCAGGAG 3232
Qy 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3233 TGGTGGATTGGCCACATGATGGAGATCTGTGTCGAAAGGCGCATTCCTCCGCTGTCAVTT 3292
Qy 1001 ValHisPheIleAlaAsp 1006
Db 3293 GTGCACCTTATYGCTGAC 3310
```

RESULT 12

```
US-09-949-002-139
; Sequence 139, Application US/09949002
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; AND USES THEREOF
; FILE REFERENCE: CL000790
```

```
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 139
; LENGTH: 3025
; TYPE: DNA
; ORGANISM: Human
; US-09-949-002-139

Alignment Scores:
Pred. No.: 5,44e-192 Length: 3025
Score: 4664.00 Matches: 893
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 2
Query Match: 88.45% Indels: 0
DB: 41 Gaps: 0
```

US-09-914-042-1 (1-1006) x US-09-949-002-139 (1-3025)

```
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 341 ATGCCGACCCAGATCTCCGTGTCGAAATTCGTGCCGAGACCCATGAGGACTACAAGGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGCGCTCCAGCTTCCACCACCGCAGCGCGAGTCCCGAAACACTGTGGCGGCGCATC 460
Qy 41 GluGluAlaLeuAspValAspAtqMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 461 GAGGAGGCTTTGGACGTGGACCGGATGCTTTTACAAATGAAGAAATCCGTGGAAGCA 520
Qy 61 IleAsnSerSerGlyLysAlaHisValLysGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCACGTGGAATAATGAAGCAGTACACCCAGCTCTGGAG 580
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 581 AAGTTTGGCGGCAACTGTGTATGCAGAGATGACCCACATTTAGGAAGTGCCTTCTGAAG 640
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTCAAAGAGGTTCACAGCAGCTTTTCAAAACCTGATTCAGATATGAAAC 700
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 701 AACATAATCTCTTCCCTTTGGACAGTTTGTGAAGGGGACCTGAAAGGAGTGAAGGG 760
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 761 GATCTGAAAAAGCCCTTTTGTATAAGCTTGGGAAGACTATGAAACAAAAAATAACCAAGATA 820
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAAGAGAAAAAGAAACACCCAGCTCCATGGGATGATTCGAGCTGAATAAAGCGGA 880
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTTGCCGAAGAGATGGAAGAGAGAGGCGCTTCTTCCAGCTACAGATGTGCGAG 940
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGluAsnLeu 220
Db 941 TATCTGCTGAAGGTCAACGAAATCAAGATTAAGGAGGAGTAGATTTACTTCAGAACTCG 1000
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValIleSer 240
Db 1001 ATCAAAATCTTTCATGCCCAATGCAATTTTTTTCAGGATGAGCTCAAGCGCTGGAAGC 1060
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTTCCATTGAAACGCTGTCTACGGATCTTTCACAGATCTTCAACGATCAACAGGCCAGGAT 1120
```

```
QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGACGTTGATACAGCTTCAGATATTTTGAATCCGATTCAGGTTGAA 1180
QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAGAGGACTCCCAATTCGTGAGAGCAGCTTATAGCTTATACATCAGCCTCAGGGA 1240
QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAGGAACATGGGACCGGAGCGGAAACGGAGCCTCTACAGAAAGAGTGACGGGATCCGA 1300
QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAAGTGTGCGCAGAAAGGAAATGTTTCAGTTAAATAATGGTTTCTGACCATATCCCATGGT 1360
QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCCCTCCCTGCAAGCTCAACCTGCTTAACCTGCGCAGGTGAAGACCAACCTT 1420
QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAGAGTGTCTTGACCTTATTTCAATGACAGACTTACCATCTTCAAGCTGAA 1480
QY 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATGAACAGGAATGTCAAAATATGGATGTCTGTGCTGCAAAATAGCAAAAGAGAGCTTTA 1540
QY 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACATGATCTTAAAGGGGATGACAAATCTCGAGAAATAATCATCTGTCGAAGACTGACA 1600
QY 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGNGATCATCTCGAAGTCGAGAGGATGACGGCAATGACGTCTGTGTGACTGTGGG 1660
QY 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCCAGATCTACATGCTTTCACCAACCTGGGCATCTGACCTGCATCGAGTGTTC 1720
QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1721 GGAATCCACCGAGAGCTGGGGGTTTCATTTATTCAGAGTGACATCTGACCTTAGATGTA 1780
QY 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGGAACATCTGAGCTGCTGCTGCCAAGAAATATTGGAAATGCAAGGCTTTAATGAGATC 1840
QY 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATCTTGCTTACCAGCTGAGGACTCAGTCAAAACCCACCCAGGCGAGCATGAT 1900
QY 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGACAC 1960
QY 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 CGGNATAACGGCGGAACCTTACAGTCTTTGCGAGGCGGTCGCAAAACGAGAGATATTTT 2020
QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATGCTCCAAAGCTTATGCTGATGCTGATGCTTACGGGAAAAATCCCATCGGCCAAC 2080
QY 581 GlyHisGluProAspGluThrAlaLeuHisIleAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACCTTGCAGTCAGATCCGTGGATCGAACCTCT 2140
QY 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrCysLys 620
Db 2141 CTTCACTTGTAGACTTTTATGTTAGAAACAGTGGGAACCTGGGATAAAACAGACAGGGGAA 2200
```

```
QY 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2201 GGCAGCAGAGCCCTGCACTACTGTGCTGACCCAGCAATGCGAGTGCCTCAAGTTGCTC 2260
QY 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2261 CTGCGGGGGAAGGCCCTCCATCAGATAGCAACGAGTCAGAGAGACTCCCGCTGGACATT 2320
QY 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCCTCAGCACGACACTGTGAGAGGCTGCTGACCCAGCCCTTATTCGAAGA 2380
QY 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTAAATCTCAGCTTTCAGTTTGAATATGATGGGACTACTCCAGAAAGACCTGATGAA 2440
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2441 AGTGATGACGACATGGATGAGAAATTTGAGGCCCTCCCAACCGCGGGAAGACCGGCC 2500
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTTCTACAGCTGGGCTCCACCCAGCTTCAGTCTAACGCTGTATCTTTGGCCAGA 2560
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCAAACTTGCACAGGAGAGCAGAGGGCTTTTCATGCCCAGCATCTTGCAGAA 2620
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2621 GAGACTTACGAGAGCCCTCTCAGTGGAGCCACCTCCCGCCCGAGCTGCAGCCCCAGC 2680
QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACCGCGCCCGCCCGCTTCTCCACGGAAATGTTGCAAAAGTTACAGACGCTCTCT 2740
QY 801 AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTGTGGAAGACAACTCTGTAAGTGTGAGCGGTGGAAGCGCGAGCATCT 2800
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCGAGCTGCTCCACCGCTGCCCTCTTTCGCGTGACATCTTACCAAT 2860
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGACGCGCCCGCCCGCTTCCCAAGAGCCCGAGCGTAAATGGAACCTTG 2920
QY 861 SerGlnProSerLysProAlaProGlyIleSerGlnIleArgProProLeuPro 880
Db 2921 AGCCAGCGAGCAGAGCTGCCCGCTGGGATCTCACAGATCAGGCCCGCCACCTCTGCC 2980
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGly 895
Db 2981 CCACAGCGCCCGACGCGCTCCCGCAGAAAGACCTGCGCGGGG 3025
```

RESULT 13

```
US-09-949-003C-925
; Sequence 925, Application US/09949003C
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH NEUROPSYCHIATRIC
; DISORDERS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000791
; CURRENT APPLICATION NUMBER: US/09/949,003C
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,446
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 74065
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 925
; LENGTH: 3025
; TYPE: DNA
```

i ORGANISM: Homo sapiens

US-09-949-003C-925

Alignment Scores:

Pred. No.: 5,44e-192 Length: 3025
 Score: 4664.00 Matches: 893
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 2
 Query Match: 88.45% Indels: 0
 DB: 41 Gaps: 0

US-09-914-042-1 (1-1006) x US-09-949-003C-925 (1-3025)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleAla 20
 Db 341 ATGCCGACACAGATCTCCGTGTGGAAATTCGTGGCCGAGACCCATGAGGACTTACAAAGCGC 400
 Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 Db 401 CCCACGGCTCCAGCTTCACCAACCGCACGGCGCAGTCCCGGNACACTGTGGCGGCATC 460
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrIleYrsMetLysLysSerValIysAla 60
 Db 461 GAGGAGGCTTTGGACGTGCGACCGGATGCTTCTTTACAAAATCAAGAAATCCGTCAAAGCA 520
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
 Db 521 ATCAACAGCTCTGGGCTGGCTCACGTGGAAATGAAGAGCAGTACCCAGGCTCTGGAG 580
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
 Db 581 AAGTTTGGCGCAACTGTGTATGCAGAGATGACCCAGATTTAGGAAGTCCGTCTCTGAAG 640
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
 Db 641 TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAAACCTGATTCAGAAATATGAAC 700
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
 Db 701 AACATAATCTCTCTCCCTTTGGACAGTTTGTCTGAAGGGGACCTGAAAGGAGTGAAGGG 760
 Qy 141 AspLeuLysLysPhePheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle 160
 Db 761 GATCTGAAAAGCCCTTTTGATAAAGCTTGAAGGACTATGAACAAAATAAACCAAGATA 820
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
 Db 821 GAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTCAAAATTAAGCGGA 880
 Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
 Db 881 GCGGAAATTCGCGAAGAGATGAAAGAGAGAGGCGCTCTTCCAGCTACAGATGTCGAG 940
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220
 Db 941 TATCTGCTGAAGTCAACGAATCAAGATTAAGAGGGAGTAGATTTACTTCAGAAATCTG 1000
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
 Db 1001 ATCAAAATCTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAGCGCTGGAAGC 1060
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
 Db 1061 CTCAAACTTCCATTGAACCGCTGTCTACGGATCTTCACAGATCAACAGCGCCCGAGAT 1120
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
 Db 1121 GAAGAAAGAGGACGCTTGATACAGCTTCAGATATTTTGAATCCGATTCGAGGTGCA 1180
 Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
 Db 1181 CAGAAAGGAGGACTCCCAATTCGTGACAGACACAGCTTATAGCTTACATCAGCCTCAGGGA 1240

Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
 Db 1241 AACAAAGGAAATGGGACCGGAGCGGAAACGGAGCCTCTACAAGAAAGAGTGACGGGATCCGA 1300
 Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
 Db 1301 AAGTGTGGCAGAAAGGAAATGTTCAAGTAAAAATGGTTTCTTGACCAATATCCATGGT 1360
 Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
 Db 1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACCTGTAACCTGCCAGGTGAAGACCAACCT 1420
 Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
 Db 1421 GAGGAAAGAAAGTCTTTTACCTTATTTTACATGACAGAACTTACCACTTTCAAGCTGAA 1480
 Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
 Db 1481 GATCAACAGGAATGTCAAAATATGATGTCTGTCTGCAAAATAGCAAAAGAAAGACTTTA 1540
 Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnIleValGlnGluLeuThr 420
 Db 1541 AACAAATGCATTTAAGGGGATGACAAATCTGAGAAATAAACATCGTCCAAAGAACTGACA 1600
 Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
 Db 1601 AAGAGATCATCTCAGAAAGTGCAGAGGATGACGGGCAATGACGCTCTGCTGTGACTGTGG 1660
 Qy 441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
 Db 1661 GCGCAGATCTCATACATGGCTTTCACCAACCTGGGCATCTCTGACCTGCATCGAGTGTTC 1720
 Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
 Db 1721 GGAATCCACGAGAGCTGGGGGTTTCATTTATCCAGAGTGCAGTCCCTGACCTTAGATGTA 1780
 Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
 Db 1781 CTGGGAACATCTGAGCTGTCTGCCNAGNATATTGGGAATGCGAGCTTTAATGAGATC 1840
 Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
 Db 1841 ATGGAATGTTGCCCTACAGCTGAGGACTCAGTCAAAACCCCAACCCAGGCGAGCATGAAT 1900
 Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaAaGlyLysHis 540
 Db 1901 GCAAGAAGGACTACATCACAGCCCAAGTACACAGAGGAGATACGCAAGGAAGAGCAC 1960
 Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
 Db 1961 GCGGATTAACGGCGGGAAGCTTCACAGCTTTTGGCGAGCCGTCNAACGAGAGATATTTT 2020
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 2021 GGAATGCTCCAAAGCTTATGCTGATGGTGTGATCTTACGGAAAAAATCCCACTGGCCAAAC 2080
 Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
 Db 2081 GGACATGAGCGGATGAACCGCCCTCACCTTCAGTCAGTCCGTGGATCGGAACCTCT 2140
 Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
 Db 2141 CTTTCACTGTAGACTTTTGTAGTTTCAGNACAGTGGGAACCTGGATTAACACAGAGGAAA 2200
 Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
 Db 2201 GGCAGCAGCCCTGCACTACTGCTGCTGACCCAGCAATGCGAGTGCCTCAAGTGTGCTC 2260
 Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
 Db 2261 CTGGGGGGAAGGCTTCATCGATGATGAACAGAGTCAAGAGAGACTCCCTCGACAT 2320
 Qy 661 AlaLysArgLeuLysHisGluHisGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680


```

2321 GCCAAGCGCTCAAGCAGCAGCACGTGTGAGGAGCTGTGACCAAGCCTTATCTGGAAGA 2380
Qy
681 PheAenSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db
2381 TTTAATCTCAGTTTCAAGTGAATATGATGGCGACTACTCCAGAGAGACCTGTGATGAA 2440
Qy
701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
Db
2441 AGTGATGACGACATGATGAGAAATGTCAGCCAGTCCCAACCGCGGGAAGACCGGCC 2500
Qy
721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db
2501 ATCAGCTTCTACAGCTGGCTGCCAACCAGCTTCAAGTCTAAGCTGTATCTTTGGCCAGA 2560
Qy
741 AspAlaAlaAsnLeuAlaGlySerGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db
2561 GATGCTGCAACCTTGGCCAGGAGGAGCAGAGGGCTTTTCATGCCAGCATCTTGCAGAT 2620
Qy
761 GluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSer 780
Db
2621 GAGACTTACGGAGCGCTCTGAGTGGCAGCCACCTCCCGCCAGCCTGCAGCCCCCAGC 2680
Qy
781 ThrThrSerAlaProProLeuProProArgAsnValGlyValGlnThrAlaSerSer 800
Db
2681 ACCACGAGCGCCCCCGCTTCTCCACGGAAATGTTGGCAAGTTCAGACAGCCTCCTCT 2740
Qy
801 AlaAenThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db
2741 GCTAACACCTGTGGAGAACAACTCTGTAGTGTGGACGGTGGAGCGCGGAGCGATCT 2800
Qy
821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db
2801 TCGTCAGATCCGCCAGCTGTCCATCCACCGCTGCCCGCTCTTCGCGTGACATCTACCAAT 2860
Qy
841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db
2861 CCCCTGACCCCGACCGCCCGCCACCGCTGTGCAAGACGCCCGAGCGTAAATGGAAGCCTG 2920
Qy
861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
Db
2921 AGCCAGCGAGCAGACCGCTGCCCGCTGGATCTCAGATCAGACGACGCCCGCCCTGCCC 2980
Qy
881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGly 895
Db
2981 CCACAGCGCGCCGCGCTCCCGCAGAGAAGCCTGCGCGCGGG 3025

RESULT 14
US-10-170-235-8660
; Sequence 8660, Application US/10170235
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HUMAN
; TITLE OF INVENTION: TRANSSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF
; FILE REFERENCE: CL001380
; CURRENT APPLICATION NUMBER: US/10/170,235
; CURRENT FILING DATE: 2003-03-17
; NUMBER OF SEQ ID NOS: 42514
; SEQ ID NO 8660
; LENGTH: 5427
; TYPE: DNA
; ORGANISM: HUMAN
US-10-170-235-8660

Alignment Scores:
Pred. No.: 1,26e-187 Length: 5427
Score: 4568.50 Matches: 893
Percent Similarity: 88.00% Conservative: 2
Best Local Similarity: 87.81% Mismatches: 3
Query Match: 86.64% Indels: 119
Db: 49 Gaps: 5

US-09-914-042-1 (1-1006) x US-10-170-235-8660 (1-5427)

```

```

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20
Db 341 ATGCGGACCCAGATCTCCGTGTGCGAAATTCGTGGCCGAGAGCCATGAGGACTACAAAGCG 400
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCAGCGCTCCAGCTTCACCAACCGCAGCGCAGTCCCGAAACACACTGTGGCGGCCATC 460
Qy 41 GluGlu-----AlaLeuAspValAspArgMet 49
Db 461 GAGGAGGTGAGCGCGCCTGCGCGCGGCTCCGCGCGCAGGCTTTGGACGTGAGCCGGATG 520
Qy 50 ValLeuTyrLysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisVal 69
Db 521 GTTCTTCAAAAATGAAGAAATCCGTGAAGCAATCAACAGACTCTGGGCTGGCTCACGCTG 580
Qy 70 GluAsnGluGluGlnTyrThrGlnAlaLeuGluLysPheGlyGlyAsnCysValCysArg 89
Db 581 GAAATGAAGAGCAGTACACCCAGCTCTGGAGAAGTTTGGCGGCAACTGTGTATGCAGA 640
Qy 90 AspAspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
Db 641 GATGCCCGAGATTTAGGAAGTGGCTTCTGAAAGTCTCAGTGTTCACAAAGAGAGTTGACA 700
Qy 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSer 129
Db 701 GCATCTTTCAAAAACCTG----- 718
Qy 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysPhePheAspLysAla 149
Db 719 -----GATCTGAAAAAGCCTTTTGTATAAGCT 745
Qy 150 TrpLysAspTyrGluThrLysIleThrLysIleGluLysGluLysLysLysLysLysLys 169
Db 746 TGGAGAGGACTATGAACAAAAATACCAAGATAGAAAAGAGAGAAAAGAACACCCCAAG 805
Qy 170 LeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLys 189
Db 806 CTCCATGGATGATTCGGACTGAAATAGCGGAGCGGAAATTCGCGAAGAGAGATGGAAGA 865
Qy 190 GluArgArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluLysLys 209
Db 866 GAGAGCGCTCTTCCAGCTACAGATGTCGAGTATCTGCTGAAGGTCAACGAAATCAAG 925
Qy 210 IleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsn 229
Db 926 ATTAAGAGGAGGTAGATTCTCAGAAATCTGATCAAAATCTTTTCATGCCCAATGCAAT 985
Qy 230 PhePheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSer 249
Db 986 TTTTTCAGGATGGACTCAAGCGCTGGAAGCCTCAAACTTCATTTGAAACGCTGTCT 1045
Qy 250 ThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgArgGlnLeuIleGlnLeu 269
Db 1046 ACGGATCTTCACAGCATCAACAGCCAGGATGAAGAAGAGAGAGAGTGTATACAGCTT 1105
Qy 270 ArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgGln 289
Db 1106 CGAGATATTTGAATATCCGATTCAGGTTGAACAGAAAGAGAGTCCCAATTCGTGAC 1165
Qy 290 SerThrAlaTyrSerLeuHisGlnProGlnGlnAsnLysGluHisGlyThrGluArgAsn 309
Db 1166 AGCAGACTTATAGCTTACATCAGCTCAGCGAAGCAAGAAACATGGGACCGAGCGGAAC 1225
Qy 310 GlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSer 329
Db 1226 GGCAGCTCTCAAGAGAGAGTACGGGTAC----- 1255
Qy 330 ValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeu 349
Db 1256 -----GCTAACCGCGCTCTCTGCAAAAGCTC 1279

```


Qy 350 AsnLeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIle 369
Db 1280 AACCTGCTTAACCTGCCAGGTGAGACCAACCTCTGAGGAGAGAGAGTGTCTTACCTTATT 1339
Qy 370 SerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMet 389
Db 1340 TCACATGACAGAACTTACCACCTTCAAGCTGAAGATGAACAGGAATGCAAAATGGATG 1399
Qy 390 SerValLeuGlnAsnSerLysGluGluAlaLeuAsnAlaPheLysGlyAspAsn 409
Db 1400 TCTGTGCTGCAAAATAGCAAGAGAAAGCTTTAAACAAATGCAATTTAAAGGGGATGACAA 1459
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
Db 1460 ACTGAGAAATAAATCGTCCAGACTGCAAAAGGAGATCATCTCAGAAAGTGAGAGG 1519
Qy 430 MetThrGlyAsnAspValCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449
Db 1520 ATGACGGGCAATGAGCTGCTGTGACGTGGGGGCCAGATCTTACATGGCTTCCACC 1579
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
Db 1580 AACCTGGGCATCTGACCTGACATGAGTGTTCGGGAATCCACCGAGAGCTGGGGTTCAT 1639
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
Db 1640 TATTCAGAGTGACGTCCGTGACCTTAGATGTACTGGGAACATCTGAGCTGCTGCTGCC 1699
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
Db 1700 AAGAATAATTGGGAATGCAAGCTTTAATCAGATCATGGAATGTGCTTACCAGCTGAGGAC 1759
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
Db 1760 TCAGTCAAAACCAACCCAGGAGCAGACATGAATGCAAGAAAGGACTACATCACAGCCAA 1819
Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
Db 1820 TACATCGAGAGAGATAGCAAGGAGAGACGCGGATTAACGCGGCGAAGCTTCACAGT 1879
Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaLysAlaAspGly 569
Db 1880 CTTTTCGAGGCGCTCAAAACAGAGATATTTTGGATTGCTTCCAAAGCTTATGCTGATGG 1939
Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
Db 1940 GTGGATCTTACGGAATAATCCACTGGCCACGAGACATGAGCCGGATGAAACGGCCCTC 1999
Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
Db 2000 CACCTTGACATCAGATCCGTGATCGAATCTCTTTCATTTGATGAGCTTTTATGTTGAG 2059
Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
Db 2060 AACAGTGGGAACCTGGATAAACAGACAGGAAAGGACGACACAGCCCTGCACCTACTGCTG 2119
Qy 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIle 649
Db 2120 CTGACCGCAATGCGGAGTGCTCAAGTTGCTTCTCGCGGGGAGAGGCTTCCATCGAGATA 2179
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 2180 GCAACAGAGTCAGAGAGACTCCGCTGGACATTTGCAAGCGCTCAAGCAGCAGACCTGT 2239
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2240 GAGGAGCTGCTGACCAAGCTTATCTGGAAGATTTAATCTCTCAGCTTCAAGTTCATAT 2299
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2300 GAATGGCGACTACTCCAGCAAGACCTGGATGAAGTGATGACGACATGATGAGAAATG 2359
Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729

Db 2360 CAGCCAGTCCCAACCGCGGAGAGACCGCCCATCAGCTTCTACCAGTGGGTCCCAAC 2419
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2420 CAGCTTCAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAACTTGGCAAGAGAG 2479
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGly 769
Db 2480 CAGAGGGCTTTTCATGCCCAGCATCTTGAGAAATGAGACTTACGGAGGCCCTCTGAGTGGC 2539
Qy 770 SerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProPro 789
Db 2540 AGCCCACTCCGCGCCAGCTGAGCCGCCAGCACCCAGCAGCCGCCCTTCTCCCA 2599
Qy 790 ArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSer 809
Db 2600 CGGAATGTTGCCAAA-----GATCCCTGACCCCGCCAGCCGCCACCC 2614
Qy 810 ValSerValAspGlyGlySerArgGlnArgSerSerSerAspProProAlaValHisPro 829
Db 2614 ----- 2614
Qy 830 ProLeuProProLeuArgValThrSerThrAsnProLeuThrProThrProProPro 849
Db 2615 -----GATCCCTGACCCCGCCAGCCGCCACCC 2644
Qy 850 ValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAlaProPro 869
Db 2645 GTTCCCAAGAGCCCGCAGCGTAATGGAAGCTTGAAGCAGCGAGCAAGCCCTGCCCGCT 2704
Qy 870 GlyIleSerGlnIleArgProProProLeuProGlnProGlnProSerArgLeuProGln 889
Db 2705 GGGATCTCACAGATCAGGCCCCACCTCTGCCCAACAGCCCGCCAGCCCTCCCGAG 2764
Qy 890 LysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909
Db 2765 AAGAAGCTGCGCGGG----- 2782
Qy 910 GlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeu 929
Db 2783 -----TTGGATCTCTTGCACCAAGCAAGCTCTGGTCTCTGCAATGCTATGTCCTG 2836
Qy 930 GlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgVal 949
Db 2837 CAGCCCCCTGACCCATGCTTAGAAGTCGCGAGCGACCAAGCTTGAAGCCTAAGCGGTG 2896
Qy 950 LysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAsp 969
Db 2897 AAGGCTCTATACTGTGTGCTGACAAACCCGATGAGCTCACCTTCTCGAGGGGAT 2956
Qy 970 ValIleValAspGlyGluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAsp 989
Db 2957 GTGATCATCTGTCGCGGGAGGAGCAGAGAGTGGTGATGGCCACATGATGGAGAT 3016
Qy 990 ProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3017 CTTGTCGCAAAAGCGCATTCCTGGTGTCATTTTGTCACCTTTATCGCTGAC 3067

RESULT 15

US-60-452-680-5750
; Sequence 5750, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5750

; LENGTH: 5427

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-452-680-5750

Alignment Scores:

Pred. No.: 2,298-187 Length: 5427
 Score: 4562.50 Matches: 892
 Percent Similarity: 87.91% Conservative: 2
 Best Local Similarity: 87.71% Mismatches: 4
 Query Match: 86.53% Indels: 119
 DB: 115 Gaps: 5

US-09-914-042-1 (1-1006) x US-60-452-680-5750 (1-5427)

QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
 DB 341 ATGCGCGACCATCTCCGCTGCGAATTCGTGGCGGAGCCATGAGGACTACAGGCG 400
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 DB 401 CCCAGGCGCTCCAGCTTCACACCGCCACGCGCGAGTGCAGAAACACTGTGGCGCCATC 460
 QY 41 GluGlu-----AlaLeuAspValAspArgMet 49
 DB 461 GAGGAGGTGAGCGCGCTCGCGCGGCTCGCGCGAGGCTTTCGAYGTGGACCGGATG 520
 QY 50 ValLeuTyrLysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisVal 69
 DB 521 GTTCTTTACAAATGAGAAATCCGTGAAGCAATCAACAGCTCTGGGCTGGCTCACGTG 580
 QY 70 GluAsnGluGluInTyrThrGlnAlaLeuGluLysPheGlyGlyAsnCysValCysArg 89
 DB 581 GAAATGAAGAGCAGCTACACCCAGGCTCTGGAGAGTTTGGCGGCAACTGTGTATGAGA 640
 QY 90 AspAspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
 DB 641 GATGACCCAGATTTAGGAAGTSCGTTCCTCAAGTTCTCAGTGTTTACAAAGGAGTGA 700
 QY 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSer 129
 DB 701 GCACITTTCAAAACCTG----- 718
 QY 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysPheAspLysAla 149
 DB 719 -----GATCTGAAAAGCCTTTTGATAAGCT 745
 QY 150 TrpLysAspTyrGluThrLysIleThrLysIleGluLysGluLysGlyGluHisAlaLys 169
 DB 746 TGGAAAGGACTATGAACCAAAATAACCAAGATAGAAAAGGAGAAAAGGAAACACGCCAAG 805
 QY 170 LeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLys 189
 DB 806 CTCATGGATGATTCGGACTGAATAAGCGGAGCGGAAATTCGCCAGAGATGGAAGAAG 865
 QY 190 GluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLys 209
 DB 866 GAGAGCGCTTCTCCAGCTACAGATGTCGAGTATCTGCTGAAGTCAACGAAATCAAG 925
 QY 210 IleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsn 229
 DB 926 ATTAATAAGGAGTAGATTAATCTCAGAATCTGATCAAAATCTTTTCATGCCCAATGCAAT 985
 QY 230 PhePheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSer 249
 DB 986 TTTTTCAGGATGGACTCAAGCCGTGGAAGCCCTCAAAACCTTCCATGAACCGCTGCT 1045
 QY 250 ThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGlnLeuIleGlnLeu 269
 DB 1046 ACGGATCTTCACCATCAACAGAGCCGAGGATGAAGAAGAGGAGGAGTGTGATACAGCTT 1105
 QY 270 ArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGlnIleArgGln 289

DB 1106 CGAGATATTTGAAATCCGCATTCGAGGTTGAACAGAAAGAGGACTCCAAATTCGTGAC 1165
 QY 290 SerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThrGluArgAsn 309
 DB 1166 AGCAGAGCTTATAGCTTACATCAGCTCAGGGAACAAGAAACATGGGACCGAGCGAAC 1225
 QY 310 GlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSer 329
 DB 1226 GGCAGGCTCTCAAGAAGAGTGACGGGTAC----- 1255
 QY 330 ValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProProAlaLysLeu 349
 DB 1256 -----GCTAACCGGCTCTCTGCAAGCTC 1279
 QY 350 AsnLeuLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuIle 369
 DB 1280 AACCTGTCTAACCTGCCAGTGAAGCAACCTCTGAGGAGAGAGTGTCTTGACCTYATT 1339
 QY 370 SerHisAspArgThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMet 389
 DB 1340 TCACATGACAGAACTTACCACTTTCAGCTGAAGATGAACAGGAATGTCAAAATATGATG 1399
 QY 390 SerValLeuGlnAsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsn 409
 DB 1400 TCTGTGCTGCAAAATAGCAAGAAAGAGCTTTAAACAATGCAATTTAAGGGGGATGACAA 1459
 QY 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
 DB 1460 ACTGGAGAAATACATCGTCCNAGAACTGACNAGAGAGATCATCTCAGAGTGCAGAGG 1519
 QY 430 MetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449
 DB 1520 ATGACGGCAATGAYGTCTGTGCTGTGGCGGCGCAGATCTTACATGGCTTTCACCC 1579
 QY 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
 DB 1580 AACCTGGGCACTCTGACCTGCATCGAGTGTTCGGAATCCACCGAGAGCTGGGGGTTCAT 1639
 QY 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
 DB 1640 TATTCAGGATGCAGTCCCTGACCTTAGATGTACTGGGAAACATCTGAGCTGTGCTGCC 1699
 QY 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
 DB 1700 AAGAATAATGGAAATGCAAGCTTTAATGAGATCATGGAATGTGCTTACCAGCTGAGGAC 1759
 QY 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
 DB 1760 TCAGTCAAAACCCACCCAGGAGGACATGAATGCAAGAAAGGACTATCATCACGCCAAC 1819
 QY 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
 DB 1820 TAYATCGAGAGAGATAGCAGAGGAGAGGAGGAGGATACCGCGGAGGCTTACAGT 1879
 QY 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
 DB 1880 CTTTTCGAGGCGGTCAAAACGAGAGATATTTTGGATTGCTCCAAGCTTATGCTGATGCT 1939
 QY 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
 DB 1940 GTGGATCTTACGGAATAAATCCCTGCGCAACCGAGATGAGCCGGATGAAACCGCCCTC 1999
 QY 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
 DB 2000 CACCTTGGCGTCAAGTCCGTGGATCGAACCTCTCTTCACATTGTAGACTTTTATTAGT 2059
 QY 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
 DB 2060 AACAGTGGAACTTGGATATAACAGACAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAG 2119
 QY 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 649
 DB 2120 CTGACCGCAATTCGGAGTGCCTCAAGTTGCTCTCGCGGGGAGGAGGAGGAGGAGGAGGAG 2179

Search completed: August 4, 2005, 16:31:38
Job time : 8888 secs

```
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 2180 CCAAYAGAGTCAGGAGAGACTCCGCTGGACATTGCCAAGCGCTCAAGCAGCAGCAGCTGT 2239
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2240 GAGGAGCTGCTGACCAAGCCTTATCTGGAAGATTAAATCTCACGTTTCAGTTGAATAT 2299
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2300 GAATGGCGACTACTCCAGCAAGACCTGGATGAAGTGTGACGACATGGATGAGAAATTG 2359
Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2360 CAGCCAGTCCCAACCGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAC 2419
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2420 CAGCTTCAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAACTTTGCCAAGGASAAG 2479
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGly 769
Db 2480 CAGAGGGCTTTTCATGCCAGCATCTTGAGAATGAGACTTACGGAGCCCTCCTGAGTGGC 2539
Qy 770 SerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProPro 789
Db 2540 AGCCCACTCCCGCCAGCTGACGCCCCAGCACACAGGCCCGCCCGCTTCTCTCA 2599
Qy 790 ArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSer 809
Db 2600 CGGAATGTTGCCAA----- 2614
Qy 810 ValSerValAspGlyGlySerArgGlnArgSerSerSerAspProProAlaValHisPro 829
Db 2614 ----- 2614
Qy 830 ProLeuProProLeuArgValThrSerThrAsnProLeuThrProThrProProPro 849
Db 2615 -----GATCCCTGACCCCGCCCGCCCGCCCGCC 2644
Qy 850 ValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLysProAlaProPro 869
Db 2645 GTTGCCAAGAGCGCCAGCGTAATGGAAGCTTGAGCCAGCGCAGCAAGCTGCCCGCGCT 2704
Qy 870 GlyIleSerGlnIleArgProProProLeuProGlnProProSerArgLeuProGln 889
Db 2705 GGGATCTCACAGATCAGGCGCCCGACCTCTGCCCGCCACAGCGCCCGCGCTCCCGCAG 2764
Qy 890 LysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909
Db 2765 ANAAGCCTGCCCGGG----- 2782
Qy 910 GlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeu 929
Db 2783 -----TTGGATCTCTGCAACGGAAGCTCTGGGTCTCTGTCCAATGCTATGTCCTG 2836
Qy 930 GlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgVal 949
Db 2837 CAGCCCGCTGCACCCATGCTAGGAAGTCGCAAGCAACCAAGTTGAAGCCTAAGCGGGTG 2896
Qy 950 LysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAsp 969
Db 2897 AAAGCGCTCTATACTGTGTGGCTGACAAACCCCGATGAGCTCACCTTCTCCGAGGGGAT 2956
Qy 970 ValIleIleValAspGlyGluAspGlnGluTyrTrpIleGlyHisIleAspGlyAsp 989
Db 2957 GTGATCATCGTGAGCGGGAGGAGGACCCAGAGATGGTGGATTGGCCACATTGATGAGAT 3016
Qy 990 ProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3017 CTTGGTCGCAAGGGCGCATTCCTCGGTGTCATTGTGTGCATTTTATYGTCTGAC 3067
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 08:58:17 ; Search time 2694 Seconds
(without alignments)
2420.646 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVFVAETHDYKA.....DGDPRKGAPFVSFVHFAD 1006

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 7297361 seqs, 3241162794 residues

Total number of hits satisfying chosen parameters: 14594722

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_b/US09914042/runat_04082005_090635_2186/app_query.fasta_1.1159
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09914042 @CGN 1.1 658 @runat_04082005_090635_2186
-NCPU=6 -ICPU=3 -NO_WMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5253	99.6	5711	21	US-10-956-157-1144	Sequence 1144, Ap
2	3006	57.0	3974	18	US-10-210-281-73	Sequence 73, Appl
3	2371.5	45.0	5582	17	US-10-062-674-1606	Sequence 1606, Ap
4	2089.5	39.6	3812	17	US-10-104-047-981	Sequence 981, Appl
5	2089.5	39.6	4050	21	US-10-490-605-1	Sequence 1, Appl
6	1801.5	34.2	3436	11	US-09-764-875-358	Sequence 37, Appl
7	1759.5	33.4	3413	11	US-09-764-875-358	Sequence 358, Appl
8	1307.5	24.8	3123	17	US-10-104-047-314	Sequence 314, Appl
9	695	13.2	563	9	US-09-815-343-1154	Sequence 1154, Ap
10	695	13.2	563	18	US-10-097-105-1154	Sequence 1154, Ap
11	571.5	10.8	6990	15	US-10-037-270-155	Sequence 155, Appl
12	571.5	10.8	6990	17	US-10-117-722-155	Sequence 155, Appl
13	570.5	10.8	2484	17	US-10-080-334-79	Sequence 79, Appl
14	561	10.6	2737	15	US-10-176-306-10	Sequence 10, Appl
15	558	10.6	5421	20	US-10-723-860-8033	Sequence 8033, Ap
16	532.5	10.1	2505	15	US-10-176-306-12	Sequence 12, Appl
17	532.5	10.1	2505	18	US-10-467-434-22	Sequence 22, Appl
18	519.5	9.9	2810	20	US-10-739-930-640	Sequence 640, Appl
19	509	9.7	6482	21	US-10-956-157-4819	Sequence 4819, Ap
20	492.5	9.3	2039	9	US-09-822-830A-256	Sequence 256, Appl
21	481	9.1	469	18	US-10-276-774-288	Sequence 288, Appl
22	479.5	9.1	399	17	US-10-542-535A-21815	Sequence 21815, A
23	479.5	9.1	399	18	US-10-085-783A-21815	Sequence 21815, A
24	453.5	8.6	2984	20	US-10-739-930-745	Sequence 745, Appl
25	441.5	8.4	2773	19	US-10-437-963-34050	Sequence 34050, A
26	429	8.1	2041	9	US-09-828-303-11	Sequence 11, Appl
27	429	8.1	2041	19	US-10-716-089-11	Sequence 11, Appl
28	428.5	8.1	442	10	US-09-318-995-27010	Sequence 27010, A
29	424	8.0	326	9	US-09-815-343-1151	Sequence 1151, Ap
30	424	8.0	326	18	US-10-097-105-1151	Sequence 19, Appl
31	416	7.9	2331	21	US-10-521-707-19	Sequence 404, Appl
32	403	7.6	409	18	US-10-376-774-404	Sequence 78764, A
33	396.5	7.5	2708	19	US-10-437-963-78764	Sequence 78764, A
34	381.5	7.4	1950	19	US-10-437-963-79456	Sequence 79456, A
35	388.5	7.4	2982	10	US-09-814-353-19609	Sequence 482, Appl
36	387.5	7.3	451	17	US-10-062-674-482	Sequence 482, Appl
37	357.5	6.8	1764	17	US-10-364-237-1383	Sequence 1383, Ap
38	340.5	6.5	3143	17	US-10-334-143-108	Sequence 108, Appl
39	336.5	6.4	3950	17	US-10-172-118-1530	Sequence 1530, Ap
40	336.5	6.4	3950	18	US-10-342-887-1530	Sequence 1530, Ap
41	333.5	6.3	3314	18	US-10-424-599-17418	Sequence 17418, A
42	330.5	6.3	2730	17	US-10-094-749-373	Sequence 373, Appl
43	329	6.2	1803	20	US-10-425-115-48371	Sequence 48371, A
44	328.5	6.2	4768	20	US-10-719-993-411	Sequence 411, Appl
45	328.5	6.2	4792	20	US-10-719-993-410	Sequence 410, Appl

ALIGNMENTS

RESULT 1

US-10-956-157-1144
; Sequence 1144, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956.157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1144
; LENGTH: 5711

```
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-956-157-1144

Alignment Scores:
Pred. No.: 0 Length: 5711
Score: 5253.00 Matches: 1003
Percent Similarity: 99.70% Conservative: 0
Best Local Similarity: 99.70% Mismatches: 3
Query Match: 99.62% Indels: 0
DB: 21 Gaps: 0

US-09-914-042-1 (1-1006) x US-10-956-157-1144 (1-5711)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleAla 20
Db 341 ATGCGGACACAGATCTCGGTGTCGGAATTCGTGGCCGAGACCCATGAGGACTACAAAGCG 400

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 401 CCCACGGCTCCAGCTTCACCAACCCGCGCGCGAGTCCCGAAACACTGTGGCGCCATC 460

Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIleAla 60
Db 461 GAGGAGGCTTTGGAGCTGGACCGGATGCTTTTACAAATGAAGATCCGTGAAGCA 520

Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 521 ATCAACAGCTCTGGCTGGCTCAGCTGCAAGCAAGTGAAGAGCAGTACACCCAGGCTCTGGAG 580

Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheIleLys 100
Db 581 AAGTTTGGCGGCAACTGTGTATGACAGATGACCCAGATTTAGGAAGTGCCTCTCGAAG 640

Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 641 TTCTCAGTGTTTACAAAGAGTTGACAGCACTTTTCAAAACCTTGATTCAGAAATATGAAC 700

Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIleGly 140
Db 701 AACATAATCTCCTTCCCTTTGACAGTTTGTCTGAAGGGGACCTGAAAGGAGTGAAAGGG 760

Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrIleLys 160
Db 761 GATCTGAAAAAGCCTTTTGATAAGCTTGGAGGACTATGAACAAAAAATAACCAAGATA 820

Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 821 GAAAAAGGAGAAAAAGGAACACGCCAAGCTCCATGGGATGATTCGGACTGAAATGAAGCGGA 880

Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 881 GCGGAAATTCGCGAAGAGATGAAAGAGAGAGGCGCTTCTTCCAGCTACAGATGTGCGAG 940

Qy 201 TyrLeuLeuLysValAsnGluIleLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 941 TATCTGCTGAAGTTCACGAATCAAGATTAAGAGGAGTAGATTTTACTTCAGAAATCTG 1000

Qy 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240
Db 1001 ATCAAAATCTTTCATGCCCAATGCAATTTTTCAGGATGGACTCAAAAGCGTGGAAAGC 1060

Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1061 CTCAAACTCTCCATTGAAACGCTGTCTACGGATCTTTCACAGATCAAAACAGGCGCCAGGAT 1120

Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1121 GAAGAAAGAGGAGCTTGATACAGCTTCGAGATATTTTGAATCCGCATTGCAGGTTGAA 1180

Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1181 CAGAAAGAGGACTCCCAAAATTCGTGAGAGCACAGCTTATAGCTTATACATCAGCCTCAGGGA 1240
```

```
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1241 AACAAAGAAACATGGGACCGGACCGGAAACGGAGCTCTACAAAGAAAGAGTGACGGGATCCGA 1300

Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1301 AAGTGTGGCAGAAAGAAAGNAATGTTTCAAGTTAAANAATGGTTTCTTGACCATATCCCATGGT 1360

Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1361 ACCGCTAACCGGCTCTCTGCAAAAGCTCAACTGCTTAACCTGCCAGGTGAAGACCAACCT 1420

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1421 GAGGAGAAGAAGTGTCTTGACCTTATTTTCAATGACAGAACTTACCACTTTTCAAGCTGAA 1480

Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1481 GATGAAACAGGNAATGTCAAATATGGATGCTGTGTCGCANAAATAGCAAAAGAAAGCTTTA 1540

Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1541 AACAAATGCATTTAAGGGGATGACAACTACTCGGAGAAATAAATCATCGTCCAAGAACTGACA 1600

Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1601 AAGGAGATCATCTCAGAAAGTGCAGAGATGACGGGCAATGACGCTCTGTGTGACTGTGGG 1660

Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1661 GCGCCAGATCTTACATGGCTTTCCACCACCTGGGCACTCTGACCTGCATCGAGTGTCC 1720

Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuLeuAspVal 480
Db 1721 GGAATCCACCAGAGCTGGGGGTTTCAATTTTCCAGGATGACGTCCTCCACCTTAGATGTA 1780

Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1781 CTGGAAACATCTGAGCTGCTGCTGCCAAGAAATATTGGGAATGCAAGCTTTTAATGAGATC 1840

Qy 501 MetGluCysLysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1841 ATGGAATGTTGCCTTACCAGCTGAGGACTCAGTCAAAACCAACCCAGCAGCAGCATGAAT 1900

Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1901 GCAAGAAAGGACTACATCACAGCCCAAGTACATCGAGAGGAGATACGCAAGGAAGAACAC 1960

Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1961 GCGGATAACCGGCGGAAGCTTTCACAGCTTTTTCGAGGCGCTCAAAACAGAGAGATATTTT 2020

Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2021 GGATTTGCTCCAGCTTATGCTGATGGTGGATCTTACGGGAAAAAATCCCACTGGCCCAAC 2080

Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2081 GGACATGAGCGGATGAAACGGCCCTCCACTTCAGTCAGTCAGATCCGTGGATCGAACCTCT 2140

Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2141 CTTTCACTTGTAGACTTTTTTAGTTTCAGAACAGTGGGAACCTGGGATTAACACAGCAGGAAA 2200

Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLysLeuLysLeu 640
Db 2201 GGCAGCACAGCCCTGCACACTACTGCTGCTGACCCGACAAATGCCGAGTGCCTCAAGTTGCTC 2260

Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGlnSerGlyGluThrProLeuAspIle 660
Db 2261 CTGCGGGGGAAGGCTCCATCGATGATAGCAAAACGAGTCAGGAGAGACTCCCGCTCGACATT 2320
```

```
QY 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2321 GCCAAGCGCTCAAGCAGACACTGTGAGGAGCTGCTGACCCAAAGCCTTATCTGGA 2380
QY 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2381 TTTAATCTCAGTTTCAGTTGAAATATGATGCGGACTACTCCAGAGACCTTGAATGA 2440
QY 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgAspGluAspArgPro 720
Db 2441 AGTGATGACGACATGGATGAGAAATGGCAGCCAGTCCCAACCGCGGGAAGACCGGCC 2500
QY 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2501 ATCAGCTCTTACACGCTGGGCTCCCAACCAAGCTTCACTTAACGCTGTATCTTTGGCC 2560
QY 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2561 GATGCTGCMAACCTTGCCNAGGAGAGCAGAGGCTTTTCATGCCAGCATCTTGCAGAA 2620
QY 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2621 GAGACTTACGAGGACCTCTGAGTGCAGCCCACTCCCGCCAGCTGCAGCCGCCAGC 2680
QY 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2681 ACCACAGCGCCCGCCCGCTTCTCCACGGAATGTTGGCAAGTTTCAGACAGCTCTCT 2740
QY 801 AlaAsnThrLeuTyrPheThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2741 GCTAACACCTTGGAAGACAAACTCTGTAAAGTGGAGCGGTGGAGCGCGGAGCGATCT 2800
QY 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
Db 2801 TCGTCAGATCCGCCAGCTGTCCATCCACCGCTGCCCTCTTCGGGTGACATCTACCAAT 2860
QY 841 ProLeuThrProThrProProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2861 CCCCTGACCCCGCCAGCGCCCGCTTGCAGAGAGCGCCAGCGGTAAATGGAAGCCTTG 2920
QY 861 SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProLeuProPro 880
Db 2921 AGCCAGCGAGAGAACCTTGCCCGCTGGGATCTCACAGATCAGGCGCCCGCCCTTGCCC 2980
QY 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2981 CCACAGCGCCCGCCAGCGCTCCCGCAGAGAGAGCTGCGCGGGGGCTGACAGTCCACC 3040
QY 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 3041 CCACTGACCAACAAAGGCCAACCGAGAGAGCCTGTGGATCTCTCTGCAACGGAAGCTCTG 3100
QY 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 3101 GGTCTCTGTCCAATGCTATGCTCTCGAGCGCCCTGCAGCCCATCTCCTAGGAAGTCGAG 3160
QY 941 AlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPro 960
Db 3161 GCAACCAAGTTGAACCTTAAGCGGTGAAGCGCTCTATTAAGTGTGGCTGACAAACCCC 3220
QY 961 AspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGlu 980
Db 3221 GATGAGCTCACCTTCTCCAGGGGATGTGATCATCGTGGAGCGGGAGGAGGACGAGGAG 3280
QY 981 TrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPhe 1000
Db 3281 TGGTCGATGGCCACATTGATGAGATCCTGTGTCGCAAGGCGCATCTCCCGGTGTCATT 3340
1001 ValHisPheIleAlaAsp 1006
3341 GTGCACTTTATCGCTGAC 3358
```

RESULT 2

```
US-10-210-281-73
; Sequence 73, Application US/10210281
; Publication No. US20040030096A1
; GENERAL INFORMATION:
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Zhong, Wei
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles E.
; APPLICANT: Ji, Weizhen
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Sciore, Paul
; APPLICANT: Stone, David J.
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Casman, Stacie
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Boldog, Ferenc L.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS
; FILE REFERENCE: 21402-416D
; CURRENT APPLICATION NUMBER: US/10/210,281
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: 60/309,501
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/310,291
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/361,775
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/361,832
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,203
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/313,201
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,702
; PRIOR FILING DATE: 2001-08-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 73
; LENGTH: 3974
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261)..(3656)
US-10-210-281-73
Alignment Scores:
Pred. No.: 1.86e-217 Length: 3974
Score: 3006.00 Matches: 626
Percent Similarity: 66.58% Conservative: 127
Best Local Similarity: 55.35% Mismatches: 234
Query Match: 57.01% Indels: 144
DB: 18 Gaps: 22
US-09-914-042-1 (1-1006) x US-10-210-281-73 (1-3974)
QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 321 ATGCGGACCAAGATCTCTGTCTCGAGTTTCATCGCGAGACCCGAGAGCTACAACTCG 380
```

Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 381 CCCACCACGTCACGCTTCCACCGCGGCTGCACAACTGCAGGAAACACCGTCACGCTGCTG 440
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 441 GAGGAGGCTCTAGGCCAAGATAGACAGCCCTTCGAAAGTGAAGAGTCTGTAAAGCA 500
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 501 ATATATAATTCTGGTCAAGATCATGTACAAATGAAGAAACTATGCAACAAGTCTTCTGAT 560
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 561 AAGTTGGGAGTAATTTTTTAAGTCGACAGACCCCGACCTTGGCAGCGCGTTTGTCAAG 620
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120
Db 621 TTTTCTACTCTTCAAGAGAACTGTCCACACTGCTGGAANAATCTGCTCCAGGTTTGAGC 680
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 681 CACAATGTGATCTTCCACCTTGGATTTCTTTTAAAGGAGACCTTAAAGGGAGTCAAAGGA 740
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 741 GATCTCAAGAGCCATTTTGACAAAGCCCTGGAAGATTTATGACAAAGTTTACAAAAATT 800
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
Db 801 GAGAAAGAGAAAGAGACACGCAAAACACATGGGATGNTCCGCACAGAGATACACGA 860
Qy 181 AlaGluIleAlaGluLysGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 861 GCTGAGATTGCGGAAGAAATGAGAAGGAAAGGCGCCTCTTTCAGCTCCAAATGTGTGAA 920
Qy 201 TyrLeuLeuLysValAsnGluLysLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 921 TATCTCAATTAAGTTAATGAAATCAAGACCAAAAGGGGTGGATCTGCTGCAGAAATCTT 980
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 981 ATAAAGTATTACCATGCACAGTGCATTTCTTCAAGATGGCTTGAAACACAGCTGATAG 1040
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1041 TTGAAACAGTACATTGAAAAATGCGCTGCTGATTTATATAATATAATAAAACAGACCAGGAT 1100
Qy 261 GluGluArgArgGlnLeuLysGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1101 GAAGAAAGAAACAGCTAACTGCACTCCGACACTTAATAAAATCCTCTTCAACTGGAT 1160
Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297
Db 1161 CAGAAAGAAATCTAGGAGAGATTCTCAGAGCGGCGAAGCA---GGATACAGATGATCG 1217
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
Db 1218 CTCAGGCGCAATAGGAATATGGCAGTGAAGAGGGGTACCTGCTCTAAAGAAAGTGCAC 1277
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 1278 GGGATCCGGAAGTATGGCAGAGAGGAAGTGTTCAGTCAAGAAATGGGATTTCTGACCATC 1337
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 1338 TCACATGCCATCTTAACAGGCAACACGCAAGTTGAACCTTCTCACCTGCCAAGTAAAA 1397
Qy 358 ThrAsnProGluLysLysCysPheAspLeuLysSerHisAspArgThrTyrHisPhe 377
Db 1398 CCTAATGCCGAGACAAAAAATCTTTTGACCTGATATACATAATAGAACATATCACATTT 1457

Qy 378 GlnAlaGluAspGluGlnGluCysGlnIlePheMetSerValLeuGlnAsnSerLysGlu 397
Db 1458 CAGCAGAAAGATGACAGGATTATGTAGCATGGATATCATGATTGTACAAATAGCAAGAA 1517
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAspThrGlyGluAsnAsnIleValGln 417
Db 1518 GAGGCCCTAACCATGGCTTCCGTGGAGAGCAGAGTGGGGAGAGACAGCCTG---GAA 1574
Qy 418 GluLeuThrLysGluIleLysSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
Db 1575 GACCTGCAAAAGCCATTATTAGGATGTCAGCGGCTCCAGCGGATCCAGCGGAATGACATTTGCTGC 1634
Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457
Db 1635 GATTGTGGCTCATCAGAACCCACCTGCTTTCACCAACACTTGGGTATTATTTGACCTGTATA 1694
Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477
Db 1695 GAATGTTCTGCATCCATAGGGAATGGGGTTTCATATTCTCGCATTCAGTCTTTGGAA 1754
Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497
Db 1755 CTAGACAAATTAGGAACCTTCTGAACCTCTGTGCGCAAGATGTAGGAAACAATAGTTT 1814
Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517
Db 1815 AATGATATTATGGAAGCAAAATTTACCCAGC---CCCTCACCAAAACCCACCCCTTCAAGT 1871
Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArg 537
Db 1872 GATATGACTGTACGAAAGAAATATATCACTGCAAGATGTAGATCATAGTTTCAAGG 1931
Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557
Db 1932 AAGACCTGTTCAACTTCACTCAGCTAAACTAAATGAATTCGTTGAGGCCATCAAAATCCAGG 1991
Qy 558 AspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577
Db 1992 GATTTACTTGTCACATAATTTCAAGTCTATGCAAGAGGGGTAGAGCTAAATGGAAACCACTG--- 2048
Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597
Db 2049 CTGGAACCTGGCAGGAGCTTGGGAGACAGCCCTTCACCTTTCGCTCCGAACCTGCAGAT 2108
Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617
Db 2109 CAGACATCTCTCCATTTGGTGTGACTTCTCTGTAACAAACCTGTGGGAACCTCGATAAGCAG 2168
Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeu 637
Db 2169 ACGCCCTGGGAACACAGATTCTACACTACTGTAGTATGTACAGTAAACCTGAGTGTGTTG 2228
Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657
Db 2229 AAGCTTTGCTCAGGACCAAGCCCACTGTGGATATAGTTAACAGGCTGGAGAACTGCC 2288
Qy 658 LeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677
Db 2289 CTAGACATAGCAAGAGAGATAAAAGTACCAGTGTGAAGATCTGCTTCCAGGCTAAA 2348
Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAsp 697
Db 2349 TCTGGAAAGTTCAATCCACAGCTCCAGTAGATATGATGGAATCTTCGACAGGAGGAG 2408
Qy 698 LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlnAsnArgArgGlu 717
Db 2409 ATAGATGAGCGGATGATGATCTCGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCA 2468
Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737
Db 2469 CCCAGACCTCAGAGCTTCTGCGCACTCTCCAGCATCTCCCCCCAGGAC-----AAG 2519
Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGlnArgAlaPheMetProSerIle 757


```
Db 2520 CTGGCACTGCCAGGATTGAGCACTCAAGGGCAAAACAGCGGCTC----- 2564
Qy 758 LeuGlnAenGluThrTyrGlyAlaLeuLeuSerGly-----SerProProAla 774
Db 2565 -----TCCTATGGAGCTTCCACCAACAGATCTTCGTTTCCACAGACACAC 2612
Qy 775 GlnProAlaAlaProSerThrSerAlaProProLeuProProArgAsnValGlyLys 794
Db 2613 TCGCCACATCACCA---ACCACGAGGCTCCCGCTCTGCCCTTAGGAACGCCGGAAA 2669
Qy 795 -----ValGlnThrAlaSerSerAlaAsn 802
Db 2670 GGTCAACTGGCCCACTTCAACACTCCTTAAGCACCCAGACCTAGTGGCAGCTCC 2729
Qy 803 ThrLeuTrpLysThrAsnSerValSerValAepGlySerArgGlnArgSerSer 822
Db 2730 ACCCTATCCNAGAGAGCGCTCTCCCCACACCCCGA---CACAGAGAACCTATCC 2786
Qy 823 AspProAlaValHisProProLeuProProLeuArg----- 835
Db 2787 GACCTCCAGCCCACTACTCTCATGGGCCCCCAAAACAAAGGCGCAGTTCTCTGGGGTAAC 2846
Qy 835 ----- 835
Db 2847 GATGGGGTCCATCTCTTCAAGTAAGACTACAAACAAGTTTGAGGACTATCCAGCAG 2906
Qy 836 -----ValThrSerThrAsnPro---LeuThrProThrProProPro 848
Db 2907 TCGAGCACAGTTCTGCAAGACTGCCCTTGCCCAAGAGTTCTCTTAACACTCTCAG 2966
Qy 849 ProValAla-----LysThr-----ProSer 855
Db 2967 AAAGTGGCACTAAGGAAACAGATCATCTCTCCCTAGACAAAGCCACCATCCCGCGAA 3026
Qy 856 ValMetGluAlaLeuSerGln-----ProSerLysProAlaPro----- 868
Db 3027 ATCTTTCAGAAATCATCAGAGTTGGCAGAGTTGCCCAAAAGCCACCACTGGAGACCTG 3086
Qy 869 -----ProGlyLysSerGlnLeuArgProProPro 878
Db 3087 CCCCCAAAGCCACAGAACTGGCCCGCCCAAGCCCAATTTGGAGATTTGCCGCTTAAGCA 3146
Qy 879 -----LeuProProGlnProPro---SerArg 886
Db 3147 GGAGAACTGCCCGCCCAACACAGCTGGGGAGCTGCCACCCCAACCCCACTCTCAGAC 3206
Qy 887 LeuProGlnLys----- 890
Db 3207 TTACCTCCCAACCAACAGATGAAGGACCTGCCCGCCCAACACACAGCTGGAGACCTGCTA 3266
Qy 891 -----LysProAlaProGlyThrAsp 897
Db 3267 GCATAATCCCAGACTGGAGATGCTCTACCCAGGCTCAGCAACCTCTGAGGTCACTG 3326
Qy 898 LysSerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSer 915
Db 3327 AAGTCACACCACTTGATCTATCCCAATGTGCAGTCCAGAGACGCCATCCAAAGCAA 3386
Qy 916 AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMet 935
Db 3387 GCATCTGAAGACTCCAACGACCTCAGCGCTACTCTGCCA---GAGACGCCGTACCACTG 3443
Qy 936 ProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCyb 955
Db 3444 CCCGAAATAATCAATACGGGGGAAAAATAAAGTAGGCGGAGTGAACCATTTATGACTGC 3503
Qy 956 ValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValLysValAspGly 975
Db 3504 CAGGACAGACAGGATGACGAGCTCACATTCATCGAGGGAGAGTGAATTCGTACACGGG 3563
Qy 976 GluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAla 995
```

```
Db 3564 GAAGAGGACCAAGGAGTGGTGGATTCGGCCACATTCGAGGACAGCCTGAAGAGGGGGTTC 3623
Qy 996 PheProValSerPheValHisPheIleAlaAsp 1006
Db 3624 TTTCCAGTGTCTTTGTTTCATATCTCTGTCGAC 3656

RESULT 3
US-10-062-674-1606
; Sequence 1606, Application US/10062674
; Publication No. US20040005559A1
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
; FILE REFERENCE: PA-0026-1 CIP
; CURRENT APPLICATION NUMBER: US/10/062,674
; PRIOR FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: US 09/625,102
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 2217
; SOFTWARE: PERL Program
; SEQ ID NO 1606
; LENGTH: 5582
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040005559A1 199975.3
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ... (5582)
; OTHER INFORMATION: a, t, c, g, or other
US-10-062-674-1606

Alignment Scores:
Pred. No.: 3,4e-169 Length: 5582
Score: 2371.50 Matches: 509
Percent Similarity: 62.40% Conservative: 105
Best Local Similarity: 51.73% Mismatches: 194
Query Match: 44.97% Indels: 176
Dbs: 17 Gaps: 22

US-09-914-042-1 (1-1006) x US-10-062-674-1606 (1-5582)
Qy 164 LysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIle 183
Db 2 AAAAGAGAGCAGCAAAACACATGGGATGATCCGACAGAGATAACAGGAGCTGAGATT 61
Qy 184 AlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeu 203
Db 62 CGCGAAGAAATGGAGAGAAAGCGCCTCTTTTCAGCTCCAAATGTGTGAATATCTCATTT 121
Qy 204 LysValAenGluIleLysLysGlyValAepLeuLeuGlnAenLeuIleLysTyr 223
Db 122 AAAGTTAATGAATCAAGACCAAAAGGTTGGATCTGCTGCAGATCTTTATAAAGTAT 181
Qy 224 PheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysPro 243
Db 182 TACCATGCACAGTGCATTTCTTTCAAGATGCTTTGAAACACAGCTGATAAGTTGAACAG 241
Qy 244 SerIleGluThrLeuSerThrAspLeuHisThrLysGlnAlaGlnAspGluGluArg 263
Db 242 TACATTGAAAAAATCGCTGCTGATTATATATAATAATAAAACAGACCCAGGATGAAGAAAG 301
Qy 264 ArgGlnLeuIleGlnLeuAtpAspIleLeuLysSerAlaLeuGlnValGluGlnLysGlu 283
Db 302 AAACAGCTAACTGCACCTCCGAGACTTATAAAATCTCTCTTCAACTGGATCAGAAAGAA 361
Qy 284 -----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 362 TCTAGGAGAGATTCTCAGAGCCGCGCAAGGA---GGATACAGCATCCATCAGCTCCAGGGC 418
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAspGlyIleArg 320
```

Db 419 AATAAGGAATATGGCAGTGAAGAAGGGGTACCTGCTAAAGAAAGTGACGGGATCCGG 478
Qy 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 479 AAAGTATGGCAGAGAGAGAGTGTTCAGTCAAGAAATGGGATTTCAACCATCTCCACATGCC 538
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 539 ACATCTACAGGCAACACCGACCTTGAACCTTCTCACCTGCCAAGTAAACCTTAATGCC 598
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 599 GAAGACAAAATAATCTTTTACCTGATATCATATAATAGAACATATCACTTTTCAGGCAGAA 658
Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 659 GATGACGAGGATATGTAGCATGGATATCAGTATTGACAAATAGCAAAAGAGAGGCCCTA 718
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 719 ACCATGGCTTCCGTGGAGAGCAGAGTGGGGAGAGAACAGCCTG---GAAGACCTGACA 775
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 776 AAAGCCATATTAGGATGTCCAGGGCTCCAGGGGAATGACATTTGCTGGATTTGTGGC 835
Qy 441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 836 TCATCAGAACCCACCTGGCTTTCAACCAACTTGGGTATTTTACCTGTATAGATGTTCT 895
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 896 GGCATCCATAGGAAATAGGGGTTCATATTCTCCGATTCAGTCTTTGGAACCTAGACAAA 955
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 956 TTAGGAATCTCTGAATCTTGTGGCCCAAGATGTAGGAACAATAGTATTTTAATGATATT 1015
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1016 ATGGAAGCAAAATTTACCAGC---CCCTCACCAAAACCCACCCCTTCAAGTGATATGACT 1072
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1073 GTACGAAAGAAATATATCACTGCAAAAGTATGTAGATCATAGTTTTCAGGAAGACCTGT 1132
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1133 TCAACTTCATCAGCTAAACTAAATGAAATTGCTTGAGGCCATCAAAATCCAGGGATTTACTT 1192
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1193 GCACATAATTCAAGTCTATGCAAGAGGGGTAGAGCTAAATGGAAACCACTG---CTGGAACCT 1249
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 1250 GGGCAGGAGCTTGGGGAACACCCCTTCACTTGCCTCCGACTCGCAGATCAGACATCT 1309
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 1310 CTCATTTTGGTTGACTTCTTGTAACAAATGTGGGAACCTGGATAAGCAGAGCGCCCTG 1369
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 1370 GGAACACAGATTTCTACACTACTGTATGTATGTACAGTAAACCTTGAGTGTTTGAAGCTTTG 1429
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 1430 CTCAGGACAGCCCACTGTGGATATAGTTAACAGGCTGGAGAACTGCCCTAGACATA 1489
Qy 661 AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 1490 GCAAGAGACTAAAGACTACCCAGTGTGAAGATCTGCTTTCCAGGCTAAATCTCGAAG 1549

Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 1550 TTCATTCACACAGTCCACAGTAGAATAIAGTGGAAATCTTCGACAGGAGGAGATAGATGAG 1609
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 1610 AGCGATGATGATCTGGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCACCCAGACCT 1669
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 1670 CAGAGCTTCTGCCACTCTCCAGCATCTCCGCCCCAGGAC-----AAGCTGGCACTG 1720
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 1721 CCAGATTTCAGCACTCAAGGGGCAAAACAGCGGCTC----- 1756
Qy 761 GluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGlnProAla 777
Db 1757 ---TCCTATGGAGCTTTCACCAACACAGATCTTCGTTTCCACACAGCAGACTCGCCCA 1813
Qy 778 AlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThr 797
Db 1814 TCACCA---ACCACGAGGCTCCCTCTGCTCTAGGAACGCGGGGAAAGGTCCAACT 1870
Qy 798 AlaSerSerAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArg 817
Db 1871 GGCCCACTTCAACACTCCCTCTAAGCACCCAGACCTCTAGT-----GGCAGCTCC 1921
Qy 818 GlnArgSerSerAspProProAlaValHisProProLeuProProLeuArgValThr 837
Db 1922 ACCCTATCCAAAGAGAGGCTCTCT-----CCCCCACCACCCGACACACAGAGAACC 1972
Qy 838 SerThrAsnProLeuThrProThrPro----- 846
Db 1973 CTATCCGACCTCCAGCCCACTACCTCATGGGCCCCCAACAAAGGCGCAGTTCTCTTGG 2032
Qy 847 -----ProProValAlaLysThrProSerValMetGluAlaLeuSer 861
Db 2033 GGTAAACGATGGGGTCCATCTCTTCAAGTAAGACTACAAACAAGTTTGAGGGACTATCC 2092
Qy 862 Gln----- 862
Db 2093 CAGCAGTCGAGCACCAGTTCTGTCAAGACAGTGCCTTGGCCCCAAGAGTTCTTCTTAAACTA 2152
Qy 863 ProSerLysProAla-----Pro 868
Db 2153 CCTCAGAAAGTGGCACTAAGGAAACACAGATCATCTCTCCTTAGACAAAGCCACCATCCG 2212
Qy 869 ProGlyIle-----SerGlnIle-----ArgProProPro--- 878
Db 2213 CCCGAAATCTTTCAGAANTCATCAAGTTGGCAGAGTTGCCACAAAGCCACCACTGGA 2272
Qy 879 ---LeuProProGlnProProSerArgLeuProGln----- 889
Db 2273 GACCTGCCCCCAAGCCACAGAACTGGCCCCCCCCCAAGCCCCCAATTTGGAGATTGGCGC 2332
Qy 889 ----- 889
Db 2333 CTAAGCCAGGAGAACTGCCCCCCAAACCAACAGCTGGGGGACCTGCCACCCAAACCCCAAC 2392
Qy 889 ----- 889
Db 2393 TCTCAGACTTACCTCCCAACCAACAGATGAAGGACCTGCCCCCCCAACCAACCAACAGCTGGA 2452
Qy 890 -----LysLysProAlaPro 894
Db 2453 GACCTGTAGCAAAATCCAGACTGGAGATGTCTCACCCCAAGGCTCAGCAACCTCTCAG 2512
Qy 895 GlyThrAspLysSerThrProLeu-----Thr 903
Db 2513 GTCACACTGAAGTACACCCCACTTGGATCTATCCCCAAATGTGTCAGTCCAGAGACGCCATC 2572

```

Qy 904 AsnLysGlyGlnPro---ArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyPro 922
Db 2573 CAAAAGGAAGCATCTGGAGGACTCCCAACGACCTCAGCGCTACT----- 2617
Qy 923 LeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGlnAlaThr 942
Db 2618 -----CTCCAGAGAGCGCCGTACCTACCTGCCAGAGAAATCAATACGGGG 2662
Qy 943 LysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGlu 962
Db 2663 AAAAAATAAGTGAGCGAGTGAAGACCATTTATGACTCCAGCGCAGACAACGATGACGAG 2722
Qy 963 LeuThrPheSerGluGlyAspVallelleValleAspGlyGluGluAspGlnGluTyrTrp 982
Db 2723 CTCATATTCAATCGAGGGAAGTGTATTATCGTCAAGGGAAGAGGACCGAGGAGTGTGG 2782
Qy 983 IleGlyHisIleAspGly---AspPro-GlyArgLysGlyAlaPheProValSerPheVa 1001
Db 2783 ATTGGCCCATCGAAGGAGNAGCCCTTGAAGAAGGAGGGGTCTTTTCAGTGTCTTTGT 2842
Qy 1001 lHisPheille 1004
Db 2843 TCATATCTGT 2852

```

RESULT 4

```

US-10-104-047-981
; Sequence 981, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 981
; LENGTH: 3812
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-981

```

Alignment Scores:

```

Pred. No.: 4,87e-148 Length: 3812
Score: 2089.50 Matches: 455
Percent Similarity: 59.20% Conservative: 153
Best Local Similarity: 44.30% Mismatches: 275
Query Match: 39.63% Indels: 145
DB: 17 Gaps: 19

```

US-09-914-042-1 (1-1006) x US-10-104-047-981 (1-3812)

```

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 47 ATGCCGAGCAGTTCAGCGTCCGCGGTCCTCGCGCGTCACCGCGGAGGACTCAGCTCC 106
Qy 21 ProThr---AlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAla 39
Db 107 CCGGCTGGGGCGCGCGCTTCGCGCCCAAGATGCCCGGTACCGAGGGCGCGCTGGCG 166
Qy 40 lIleGluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLys 59
Db 167 CGGGAGGAGATCTTGGAGGAGACCAACCATCTCGCAGAGAAATAAGAGGCTGTGGCG 226
Qy 60 AlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeu 79
Db 227 GCATTCATAGCTCCGGCTTGGCCATCTGGAGATGAAGACGAGTACCGAGAGCGCGTG 286
Qy 80 GluLysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeu 99
Db 287 GAATCTTTAGGCAACAGCACCCTGTGCCAGAACAGCCATGAGCTGTCCACAGGCTTCTTA 346

```

```

Qy 100 LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMet 119
Db 347 AACTTGGCGCGTGTCCCGGAGGTTGCTCGCTCTTCAAGAACTGATTTCAGAACTTG 406
Qy 120 AsnAsnIlelleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLys 139
Db 407 AACAACTATGCTCTTTCCCTCGGACAGTCTGATGAAGGGGCGAGCTGAGGACGGTGA 466
Qy 140 GlyAspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLys 159
Db 467 CAGATTTCAAAAAAGCAGCTGAGAGGATGGAAGGACTATGAAGCCAAATGGCCAG 526
Qy 160 lIleGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSer 179
Db 527 CTGAGAGAGGAG---CGCGATCGGCGCAGGCTGACAGGAGGATC----- 568
Qy 180 GlyAlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCys 199
Db 569 CCTGGGAGGTGGGCCCAAGACATGCAGAGAGGCGCGCATCTTCCAGCTGCACATGTGT 628
Qy 200 GluTyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsn 219
Db 629 GAGTATCTGTCAAAAGCGGGGAGAGCCAGATGAAGCAAGTCTCTGATCTTCTTCAGAGC 688
Qy 220 LeuIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu 239
Db 689 CTCATCAAGTCTTCCCGCCAGCACAACTTTTCCAAGATGGCTGGAAGGCTGCCCCAG 748
Qy 240 SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln 259
Db 749 AGCCTGTTCCCTTTCATCGAAGCTCGCGCCCTCAGTACATGCATCCATCAGGCCAG 808
Qy 260 AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279
Db 809 GAGGACGAGCTACAGAAGCTGACCCAGCTCGGAGACTCCCTCCAGGAGACACTGAGCTT 868
Qy 280 GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln 297
Db 869 GAGAGCAGAGAGGAACACCTTGAGCGGGAAGAACTCAGGATGTGGCTATAGCATCCACCAG 928
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
Db 929 CACCAAGGCAACAAGCAGTCTTGGGACGAGAGAGTGGGCTTCTATACAAGAAAGTGAC 988
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 989 GGAATTGAGAGTCTGCGAGAAAGGAAGTGTGGAGTCAAGTATGGCTGCTGACCATC 1048
Qy 338 SerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 1049 TCACACAGCAGATAAACCGGCCCGCGTGAAGCTGACCCCTGCTGACGTGCCAAGTAGG 1108
Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
Db 1109 CCAAAACCTCAGGAGAAAAAGTGTTCGACCTGGTGCCCAACACCGGAGCTGACCATTT 1168
Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
Db 1169 CAGGACAGGACGAGCAGAGTGTGAGCGGTGGGTGTGAGTGTTCAGTGTTCAGAAACAGAG 1228
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAsp-----AspAsn 409
Db 1229 GAAGCCCTGAGCAGCGCTTCTCGGGGAGGCCAGCGCTGGCCCGGGGTCTCGGGGGTCC 1288
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
Db 1289 GCCCGCCATGATGGGAGCGCAGACCTCAAAAGTGTGCTCATCGCGGAGGTGAAGAGC 1348
Qy 430 MetThrGlyAsnAspValCysCysAspGlyAlaProAspProThrTrpLeuSerThr 449
Db 1349 AGGCTGGGAATACCCAGTGTGCGACTGCGGGGCTGCAGAGCCCAACCGTGTCTCAGCAC 1408

```

```
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
Db 1409 AACCTGGCGTGTCTCACCCTGATCGTTCGGCGTCCACCGGAACTGGCGTGGCG 1468
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
Db 1469 TTTTCGGCATGTCAGTCACTCACCCTTGACCTGCTGGGCCCTCGAGTGTGTGTCGCC 1528
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
Db 1529 TTGAACATGGGAAACACAGAGCTTCAATGAGGTTCATGGAGGCCACCTACCTCACACGGC 1588
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
Db 1589 GGCCCTAAACCTCAGCTGAGAGTGACATGGCCACCCGAGGACTACATTATGGCCAAAG 1648
Qy 530 TyrIleGluArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLysHisSer 549
Db 1649 TATGTGGAGCATAGTTTGCACGCCG-----TGCACACCTGAGCCTCAGCGA 1696
Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
Db 1697 CTCTGGACAGCATTTGCAACAGGAGCCTCTGCTGCTACTGGAGGCCCTTTGCCAATGGG 1756
Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
Db 1757 CAGGACTTGGACAGCCCGCTGCCAGGCTGATGCACAGGCACCTGGAAGAACTCGTCTTG 1816
Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
Db 1817 CATTGGCTGTCAAAAGTCCCAACAGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1876
Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
Db 1877 AACGGTGTCTACCTGGATGCCAAGCTGCTGCACGGGAAACAGCGCTCTGCATCAGCAGCA 1936
Qy 630 LeuThrAspAsnAlaGluCysLeuLeuLeuLeuArgGlyLysAlaSerIleGluIle 649
Db 1937 CTCTAACACCGCCGCTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1996
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 1997 GTAAATGAAGCAGCGGAGACAGCTCTGCACATAGCCAGGAAGACCAACCAAGAGGTGT 2056
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2057 GAGGAGCTGTGGAGAGCCCGCAGCGGGGACCTTTGCTCTTCCCTCTACATGTGGACTAC 2116
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2117 TCCTGGGTAAATTCACAGAGCTGGCTCTGACAGTGAGGAGATGAGGAAGAGAG-- 2173
Qy 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2173 ----- 2173
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2174 -----CGCTGCTTGTCTGAAGCTCCCGGCCAGGCT 2203
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
Db 2204 CACTGGGCCAGTGGGAGCTGGACATCAGCAACAGACTATGAGACTGTGCCAGCCTG 2263
Qy 769 GlySerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro 788
Db 2264 GGAGCA-----GCCACCCCTCAGGCGAGAGTGAGGAGCTGCTCCCGCCCTTGCCA 2314
Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuThrLysThrAsn 808
Db 2315 GTCAAAAC-----TCTTCTCGGACTTTG----- 2338
Qy 809 SerValSerValAspGlyGlySerArgGlnArgSer-----Ser-SerAspProPr 825
```

```
Db 2339 -----GTCCAAAGGGTGTGCAAGACATCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2389
Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2390 CTGAGTTTCAGAGGCCCTCGAGACCCCTGAGAGCCTGGGCAGTCCAG----- 2435
Qy 845 rProProProProValalalysThrProSerValMetGluAlaLeuSerGlnProSerLy 865
Db 2436 -CCTCCTCTCTCAGTCTGATGAGCCCT-----TGGAACTTGGGGATCCCGACCA 2485
Qy 865 sProAlaProProGlyIleSerGlnIleArgProProProLeuProGlnProProSe 885
Db 2486 GCCCACCACCACTCTGAAGAGGGCTCCGAGAGCCCCAGGACCTCCAGACCAGCCTG 2545
Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLy 905
Db 2546 ACAT-----CCGGGACCACTTCCGAGATGTACC----- 2576
Qy 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2577 -----TCCCGCTCAGAT---TCAGCTCCGAGAGCCTCGCTCTATCGCGG 2620
Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 2621 GGGGGCGGAGCCCTGAAGATGTCCTCAGCCAGG-CAGCTCTGCCAGAGGAAGCT 2679
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2680 GCCG----- 2683
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980
Db 2683 ----- 2683
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2684 -----GTTGGCATCCTGAGGAGATGGCTCAGGACTGGGAGTCTCCAGCAGTTC 2736
Qy 1000 eValHisPheIleAlaAsp 1006
Db 2737 TGTGCAACTTTTGCAGAC 2755
RESULT 5
US-10-490-605-1
; Sequence 1, Application US/10490605
; Publication No. US20050019768A1
; GENERAL INFORMATION:
; APPLICANT: JAPAN AS REPRESENTED BY THE PRESIDENT OF THE UNIVERSITY OF TOKYO
; APPLICANT: ONCOTHERAPY SCIENCE, INC.
; TITLE OF INVENTION: HEPATOCELLULAR CARCINOMA-RELATED GENES AND POLYPEPTIDES, AND METH
; TITLE OF INVENTION: FOR DETECTING HEPATOCELLULAR CARCINOMAS
; FILE REFERENCE: 25371-029NATL/SEN-A0121P-US
; CURRENT APPLICATION NUMBER: US/10/490,605
; CURRENT FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: USN 60/324,261
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: CA
; PRIOR FILING DATE: 2002-08-23
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4050
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-490-605-1
Alignment Scores:
Pred. No.: 5,26e-148 Length: 4050
Score: 2089.50 Matches: 455
Percent Similarity: 59.20% Conservative: 153
Best Local Similarity: 44.30% Mismatches: 275
Query Match: 39.63% Indels: 145
```



```
Db 2116 TCCTGGGTAATTTCCACAGAGCCTGGCTCTGACATGAGGAGGATGAGGAAGAAG--- 2172
Qy 710 GlnProSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2172 -----
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2173 -----
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
Db 2203 CACTGGGCGCAGTGGGAGGCTGGACATCAGCAACAGACCTATGACAGCTGCGCCAGCCTG 2262
Qy 769 GlySerProProAlaGlnProAlaGlnProAlaLysProSerThrThrSerAlaProProLeuPro 788
Db 2263 GGAGCA-----GCCACCCCTCAGGCGGAGAGTGAGGACTGTGCCCCGCCCTTGCCA 2313
Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsn 808
Db 2314 GTCAANAAC-----TCTTCTCGGACTTTG----- 2337
Qy 809 SerValSerValAspGlyGlySerArgGlnArgSer-----Ser-SerAspProPr 825
Db 2338 -----GTCCAAGGGTGTCAAGACATGCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2388
Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2389 CTGAGTTTCAGAGGCGCCCTGAGACCCCTGAGAGCCTGGCGAGTCCAG----- 2434
Qy 845 rProProProProAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2435 -CCTCTCTCTCAGTCTGATGAGCCCT-----TGGAACTTGGGGATCCAGCCAA 2484
Qy 865 sProAlaProProGlyIleSerGlnIleArgProProProLeuProProGlnProProSe 885
Db 2485 GCCCCACCACTCTGAAGAGGGCTCCGAGAGCCCCCAGGACCTCCAGACCCAGCCTG 2544
Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905
Db 2545 ACAT-----CCGGGACCACTTCCGAGAGTGATACC----- 2575
Qy 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2576 -----TCCCGTTCAGAT---TCAGTCCGAGAGCCTCGCTCTATCGGCGG 2619
Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 2620 GGGGCGCGGAGCCCTGAAGATGGTCCCTCAGCCAGG-CAGCCTCTGCCCAGAAGGAAGT 2678
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2679 GCCG----- 2682
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnG1 980
Db 2682 ----- 2682
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2683 -----GTTGGCATCACTGAAGGAGATGGCTCAAGGAGCTGGGAGTCTCCCAAGCAAGTTC 2735
Qy 1000 eValHisPheIleAlaAsp 1006
Db 2736 TGTGCAACTTTTGCAGAGC 2754
```

RESULT 6

```
US-09-764-875-37
; Sequence 37, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
```

```
; FILE REFERENCE: PUZ02
; CURRENT APPLICATION NUMBER: US/09/764.875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 3436
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-37

Alignment Scores:
Pred. No.: 2,82e-126 Length: 3436
Score: 1801.50 Matches: 346
Percent Similarity: 87.85% Conservative: 1
Best Local Similarity: 87.59% Mismatches: 3
Query Match: 34.16% Indels: 45
Db: 11 Gaps: 1

US-09-914-042-1 (1-1006) x US-09-764-875-37 (1-3436)

Qy 612 GlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThr 631
Db 1 GGGAACTCTGGATAAACAGACAGAGGGAAGGCAGCAGCCCTGCGACTACTGCTGCTGACC 60
Qy 632 AspAsnAlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsn 651
Db 61 GACAATCCGAGTCCCTCAAGTTGCTCTCGGGGGGAAGGCCTCCATCGAGATAGCAAAAC 120
Qy 652 GluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGlu 671
Db 121 GAGTCAGAGAGAGACTCGCTGGAATTCGCCAAGCGCCTCAAGCAGCAGCACTGTGAGGAG 180
Qy 672 LeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrp 691
Db 181 CTGTGACCCCAAGCCTTATCTGGAAGATTTAATTCACGTTTACGTTGAATATGAATGG 240
Qy 692 ArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPro 711
Db 241 CGACTACTCCACGAAAGACCTGGATGAAGATGATGACGACATGGATGAGAAATTTGAGCGCC 300
Qy 712 SerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeu 731
Db 301 AGTCCCAACCGCGGGAAGACCGGCCCATCAGCTTCTACCACTGGGCTCCCAACAGCTT 360
Qy 732 GlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArg 751
Db 361 CAGTCTAACGCTGTATCTTTGGCCAGAGATGCTGCAAACTTTGCCAAGGAGAGCAGAGG 420
Qy 752 AlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPro 771
Db 421 GCTTTTCATGCCCAGCATCTTCGAGAATGAGACTTACGGAGGCCCTCTCAGTGGAGCCCA 480
Qy 772 ProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsn 791
Db 481 CCTCCGCCCAAGCCTCGAGCCCCCAGCAGCAGCCCGCCCCCGCTTCTCCACAGGAAT 540
Qy 792 ValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSer 811
Db 541 GTTGGCAAA----- 549
Qy 812 ValAspGlyGlySerArgGlnArgSerSerAspProProAlaValHisProProLeu 831
Db 549 ----- 549
Qy 832 ProProLeuArgValThrSerThrAsnProLeuThrProThrProProProValAla 851
Db 550 -----GATCCCTGACCCCCCAGCGCGMSCCCACCGCTTGC 585
Qy 852 LysThrProSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIle 871
Db 586 AAGACGCCCGAGCGTAATGGAAGCCTTGAGCCAGCGCAGCAAGCCTGCCCCCGCTGGGATC 645
```

QY 872 SerGlnIleArgProProLeuProProGlnProProSerArgLeuProGlnIleLys 891
 DB 646 TCACAGATCAGCGCCCACTCTGCCCCACAGCGCCGCTCCCGCAGAGAAG 705
 QY 892 ProAlaProGlyThrAspLeuSerThrProLeuThrAsnLysGlnProArgPro 911
 DB 706 CTGTGGCGGGGGCTGACAGTCCACCCCTGACCAACAAAGGCCAACCCGAGAGACCT 765
 QY 912 ValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnPro 931
 DB 766 GTGGATCTCTCTGCAACGGAAGCTCTGGTCTCTGTCTCAATGCTATGCTCTGCAGGCC 825
 QY 932 ProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAla 951
 DB 826 CTGTCAACCCATGCTAGGAGTCCGAGCAACCAAGTTGAAGCCCTTAAGCGGGTGAAGCG 885
 QY 952 LeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIle 971
 DB 886 CTCTATAACTGTGTGGCTGACAAACCCCGATGACCTCTCTCCGAGGGGGATGTGATC 945
 QY 972 IleValAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGly 991
 DB 946 ATCGTGGACGGGAGGAGCAGCAGAGTGTGGATTGGCCACATTTGATGGAGATCCTGCT 1005
 QY 992 ArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
 DB 1006 CGCAAGCGCATTCCTCGGTGCTATTTGTGCACCTTTATCGCTGAC 1050

RESULT 7

US-09-764-875-358
 ; Sequence 358, Application US/09764875
 ; Publication No. US20040018969A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P0202
 ; CURRENT APPLICATION NUMBER: US/09/764,875
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1249
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 358
 ; LENGTH: 3413
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-875-358

Alignment Scores:

Pred. No.: 4,266-123 Length: 3413
 Score: 1759.50 Matches: 338
 Percent Similarity: 86.48% Conservative: 1
 Best Local Similarity: 86.22% Mismatches: 8
 Query Match: 33.37% Indels: 45
 DB: 11 Gaps: 1

US-09-914-042-1 (1-1006) x US-09-764-875-358 (1-3413)

QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634
 DB 8 GATTAACAGACAGGGAAGGAGCAGCAGCCCTGCACTACTGCTGCTCACCAGCAATGCC 67
 QY 635 GluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654
 DB 68 GAGTGCCTCAAGTGTCTCTGGGGGGAAGGCTCCATCGAGATAGCAACGAGTCAGGA 127
 QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuThr 674
 DB 128 GAGACTCCGCTGGACATTGCCAAGCGCTCAAGCAGCAGCAGCTGTGAGGAGTGTGACC 187
 QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeu 694
 DB 188 CAAGCTWTATYTGGGAAGATTTAAATCTCACGTTCACTGTTGAATGAATGATGGCGACTACTC 247

QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714
 DB 248 CACGAAGACCTGGATGAAAGTGATGACGACATGATGAGAAATTGACGCCAGTCCCAAC 307
 QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734
 DB 308 CGGCGGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAAACAGCTTCAGTCTAAC 367
 QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754
 DB 368 GCTGTATCTTTGGCCAGAGATGCTGCAAACTTGGCAAGGAGAGAGCAGAGGGCTTCATG 427
 QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774
 DB 428 CCCAGCATCTTGCAGAAATGAGACTTACGAGCCCTCTCTAGTGGCAGCCACCTCCGCC 487
 QY 775 GlnProAlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794
 DB 488 CAGCTCGAGCCCCCAGCAGCACCAGCGGCCCTCCCTCCACKGAATGTTGGCAAA 547
 QY 795 ValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGly 814
 DB 547 ----- 547
 QY 815 GlySerArgGlnArgSerSerAspProProAlaValHisProProLeuProProLeu 834
 DB 547 ----- 547
 QY 835 ArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysThrPro 854
 DB 548 -----GATCCCTTACCCCTCCACCGCCGMSGCCACCCGTTGCCAAGACGCC 592
 QY 855 SerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyIleSerGlnIle 874
 DB 593 AGCGTATGGAGGCTTGAGCCAGCCGAGCAAGCTGCCCGCCCTGGGATTTCCACAGATC 652
 QY 875 ArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPro 894
 DB 653 AGGCCCCCACCCTCTGCCCCACAGCCGCCAGCCCTCCCGCAGAGAGAGCTGCGCGG 712
 QY 895 GlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLeu 914
 DB 713 GGGGCTGACAAAGTCCACCCCACTGACCAACAAAGGCCAACCGAGAGGACCTGTGGATCTC 772
 QY 915 SerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPro 934
 DB 773 TCTGCAACGGAGCTCTGGGTCTCTGTCCAAATGCTRTGCTCTGACGCCCTCCCTGACCC 832
 QY 935 MetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsn 954
 DB 833 ATGCTTAGGAAGTCGACAGGCAACCAAGTTGAAGCTTAAGCGGTGAAAGCGCTCTATAAC 892
 QY 955 CysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAsp 974
 DB 893 TGTGTGCTGACAAACCCCGATGACCTCACCCTCTCCGAGGGGGATGTGATCATCTGGAC 952
 QY 975 GlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgLysGly 994
 DB 953 GGGGAGGAGGAGGAGGAGTGTGATTTGGCCACATTTGATGGAGATCTCTGGTGGCAAGGC 1012
 QY 995 AlaPheProValSerPheValHisPheIleAlaAsp 1006
 DB 1013 GCATTCCTCGGTGCTATTTGTGCACCTTTATCGCTGAC 1048

RESULT 8

US-10-104-047-314
 ; Sequence 314, Application US/10104047
 ; Publication No. US20030236392A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
 ; FILE REFERENCE: H1-A0105


```
Db 362 TCAGAAAGAGACTACATCAGCCAGTACATCAGAGGANGATACCCCAAGGAAG 421
Qy 540 His-AlaAspAsnAlaLaLysLeuHisSerLeuGluAlaValLysThrArgAsp11 559
Db 422 CACCCCGGATACGCGCGCGA----- 443
Qy 559 ePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLys 575
Db 444 ----AANCTTTCCAGTCTTTTGNNAAGGCGNGTCCAAACCNANAAG 488

RESULT 10
US-10-097-105-1154
; Sequence 1154, Application US/10097105
; Publication No. US20040037842A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: King, Gordon E.
; APPLICANT: Segrist, Heather
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.504C1
; CURRENT APPLICATION NUMBER: US/10/097,105
; CURRENT FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 1562
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1154
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 403..446, 464, 465, 472, 483, 485, 499, 539, 554
; OTHER INFORMATION: n = A,T,C or G
US-10-097-105-1154

Alignment Scores:
Pred. No.: 1,95e-43 Length: 563
Score: 695.00 Matches: 144
Percent Similarity: 82.12% Conservative: 3
Best Local Similarity: 80.45% Mismatches: 14
Query Match: 13.18% Indels: 18
DB: 18 Gaps: 1

US-09-914-042-1 (1-1006) x US-10-097-105-1154 (1-563)

Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 2 AACATGCAITTAAGGGGGATGACAATATCTGGAGAAAAATAACATGTCACGAAGACTGACA 61
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysAspCysGly 440
Db 62 AAGGAGATCACTCAGAAGTCTGAGGATGACGGCAATGACGTCTGCTGACTGTGGG 121
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 122 GCGCCAGATCCTACATGCTTTCACCAACCTGGGCATCCTGACCTGATCGAGTGTCC 181
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 182 GGAATCCACCGAGAGCTGGGGGTTTCATTATTCCAGGATGACGCTCCCTGACCTTAGATGTA 241
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 242 CTGGGAACATCTGAGCTGCTGCCCAAGAAATATTGGGAATGCAAGGCTTTAATGAGATC 301
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer-AspMetAs 520
Db 302 ATGGATGTTGCCTACAGCTGAGGACTCAGTCAACCCCAACCCAGGCGGACATGAA 361
Qy 520 nAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg-ArgTyr-AlaArgLysLys 539
```

```
Db 362 TCAGAAAGAGACTACATCAGCCAGTACATCAGAGGANGATACCCCAAGGAAG 421
Qy 540 His-AlaAspAsnAlaLaLysLeuHisSerLeuGluAlaValLysThrArgAsp11 559
Db 422 CACCCCGGATACGCGCGCGA----- 443
Qy 559 ePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLys 575
Db 444 ----AANCTTTCCAGTCTTTTGNNAAGGCGNGTCCAAACCNANAAG 488

RESULT 11
US-10-037-270-155
; Sequence 155, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghaast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt FL_genes Version 1.0
; SEQ ID NO 155
; LENGTH: 6990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (61)..(2397)
US-10-037-270-155

Alignment Scores:
Pred. No.: 1.09e-32 Length: 6990
Score: 571.50 Matches: 196
Percent Similarity: 41.68% Conservative: 137
Best Local Similarity: 24.53% Mismatches: 305
Query Match: 10.84% Indels: 161
DB: 15 Gaps: 25

US-09-914-042-1 (1-1006) x US-10-037-270-155 (1-6990)

Qy 38 AlaAlaIleGluGluAlaLeu---AspValAspArgMetValLeuTyrLysMetLysLys 56
Db 115 GCAGCTTTGGGAAGAGTAGAAGGTGTGCGCAATTTGGAACATA---AAACTTGATGAG 171
Qy 57 SerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThr 76
Db 172 CTTGTGAAACTT-----GGGATTGCAATGATTGATCTGGAAGAGCCTTTTGT 219
Qy 77 GlnAlaLeuGluLysPheGlyGlyAsnCysValCysArgAsp----- 90
```

Db 220 GTTGCAATAAACAAGTTCATGAATGGG-----ATTGAGAGACCTGGCCAGTATTCTAGT 273
 Qy 91 ---AspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
 Db 274 AATGATGCTGCTGTGAGACAAGTTTGACCAAGTTTCTGCAGCTCTTCAAGAAATGATA 333
 Qy 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnAenIleIleSerPheProLeuAspSer 129
 Db 334 AATTTTCACACATCTCTGTGCAAACTCAGAGATCAATTTAAGCCACAGCTTCAGTAAC 393
 Qy 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysPheAspLysAla 149
 Db 394 TTGTGTTAAGAAGATCTTAGAAAATTCAAA---GATGCCAAGAAGCAATTTGAAAAGTC 450
 Qy 150 TrpLysAspTyrGluThrLysIleThrLys-----IleGluLysGluLysLysGlu 166
 Db 451 AGTGAAGAAAAGAAAATGCGTTAGTAAAAAATGCCCAAGTACAAAGAAAACAAACAT 510
 Qy 167 HisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGlu 186
 Db 511 GAAGTTGAA-----GAAGCCACCACCATTT 534
 Qy 187 MetGluLysGluArgArgPhePheGlnLeuMetCysGluTyrLeuLeuLysValAsn 206
 Db 535 CTGACAGCAACAAGAAAATGTTCCGACACATAGCCCTCGATTATGTGCTTCAGATTAAT 594
 Qy 207 GluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAla 226
 Db 595 GTTCTTCAATCAAAAAGAGATCAGAAAATCTTAAATCAATGTTGTCTATTATGTATGCC 654
 Qy 227 GlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGlu 246
 Db 655 CATTGGCTCTCTTCATCAGGATATGATCTGTTAGTGAACCTTGGACCTTACATGAAG 714
 Qy 247 ThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgArgGlnLeu 266
 Db 715 GATCTTGTGTGCACAGTGTGATCGACTGTTGTGGATGCGAGCAAGAGAGAAAAGATAATG 774
 Qy 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluLysGluAspSerGln 286
 Db 775 GAGCAAAAACATTC-----ACCATTCAACAAAAGGATTTCTCCAGT 816
 Qy 287 IleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306
 Db 817 GATGATTTCAAGTTAGATATAACGTA-----GATGCTGCAAAATGGCATA 861
 Qy 307 GluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArg 326
 Db 862 GTTATGGAAGGATATCTGTTCAACAGCGCCAGCAATGCTTCAAAAATTTGGAACAGGCGC 921
 Qy 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346
 Db 922 TGGTTTTCATACAGAAATACAGTTGGTTTACCAGAAAAAATTTAAGGATAATCCGACT 981
 Qy 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGluLys 363
 Db 982 GTGTAGTTGAAGACCTCAGGCTTTGCACATGAAAAACATTTGGAAGACATAGAGCGCAGA 1041
 Qy 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383
 Db 1042 TTCTGCTTTGAGTGTCTCGCCAAACAAAAAGTTGCTGCTCCAGCGCAGATTTCCGAAAAG 1101
 Qy 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403
 Db 1102 CTGCGCCAGGATGGAATTAAGCTGTTTCCAGCCAGTATTGCTTACTGCTTATAGAGAG--- 1158
 Qy 404 PheLysGlyAspAsp-----AsnThrGlyGlu--- 412
 Db 1159 ---AAGGTGATGAATCAGAGAGCTGGATAAGAAATCATCTCCATCCACAGGAAGCCTA 1215
 Qy 413 -----AsnAsnIleValGlnGluLeuThrLys-----GluIleIleSerGluVal 427
 Db 1216 GATTCTGGAATAGTCCAAAGAGAAATTTTGAAGAGAGAAAGTGCGCTTCAGCGGGTCT 1275

Qy 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447
 Db 1276 CAGTGTATCCCTCGCAATGCCAGCTGTGTGACGTGTGGCTCGGAGATCCACGGTGGGCC 1335
 Qy 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467
 Db 1336 AGCATCAACCTGGGCATCACCCTGTGTATCAGTGTCTCCGAATTCACCGGAGCTTGGG 1395
 Qy 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487
 Db 1396 GTTCATTTTCAAAAGTACGATCTTTAACTTTAGACACCTGGAGCCAGAACTTTTAAAG 1455
 Qy 488 LeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAla 507
 Db 1456 CTTATGTGTGAGTTGGGGAATGATTTATAAATCGAGTTTATGAAGCTAATGTGGAAAA 1515
 Qy 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527
 Db 1516 ATGGGAATAAAGAAACCCCAACCAAGGA---CAAAGACAGGAGAGGACATATATCAGA 1572
 Qy 528 AlaLysTyrIleGluArgArgTyrAla----- 536
 Db 1573 GCAAAATATGTGGAGAGGAAAATTTGTGGATAAATATTCTATATCATTTATCATCTCCTGAG 1632
 Qy 537 ---ArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCys----- 551
 Db 1633 CAGCAAAAAAAGTTTGTCTTAAAGTTCTGAAGAAAGAGGCTGAGCATTTCTAAATTT 1692
 Qy 552 -----GluAlaValLysThrArgAspIlePhe 560
 Db 1693 GGGCCAGGGGACCAAGTCAGAGCATCTGCCAAAGTTTCAGTCAGAAGTAATGAC---AGT 1749
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 1750 GGAATTCAGCAGAGCTCTGATGATGGAAGAAATCTTTACCTCCAGGCTGTGAGCCAAT 1809
 Qy 581 Gly---HisGluProAsp----- 585
 Db 1810 AGTTTATATGAGCTGAAGGAGAAAGCAAGATTCTTCTATGTTCTTGACTCGAAACAT 1869
 Qy 585 ----- 585
 Db 1870 CTTAATCCAGGACTTCAGCTTTATAGGCGCTCATATGAANAACCTTCTCTAAATGGCT 1929
 Qy 586 GluThrAlaLeuHisLeuAla----- 592
 Db 1930 GAGCTTTGGCTCATGTGTCAGACGTGAACCTGGGCCAAATTCAGAGGAAAAACAAAGCGACA 1989
 Qy 593 -----ValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsn 610
 Db 1990 CCATTTATTCAGGCTGATTAGGGGGCTCTTTTGTGAGCTGTGAGTTCTCTCTACAGAA 2049
 Qy 611 SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeu 630
 Db 2050 GGTCTAATGTCAACCAAGAGATGTCCAAGGGCGGGACCATTTGCACCATGCCACCGTC 2109
 Qy 631 ThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAla 650
 Db 2110 TTAGGGCACACAGGGCAGGTATGTTTATTCCTAAACAGAGGTGCCAATCAACATGCCACT 2169
 Qy 651 AsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlu 670
 Db 2170 GATGAAGAGGAAAGACCTTTTGGATACATCTGTGGAGCAGGCAATGTGTATATAGTC 2229
 Qy 671 GluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGlu 690
 Db 2230 ACCTTGTTACGTTTAGCA----- 2247
 Qy 691 TrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
 Db 2248 -----AGAATGAATGAAGATGCGGGAATCAGAGGACGCTTATGGA-----CAG 2292

QY 711 ProSerGluAsnArgArgGluAspArgProIleSerPheThrGlnLeuGlySerAsn 729
 Db 2293 CCAGGTGTAAGAACTTATCAGGACATATTCGTGATTTTCCCAATATGGCATCAAT 2349

RESULT 12

US-10-117-722-155
 ; Sequence 155, Application US/10117722
 ; Publication No. US20030219744A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Y. Tom
 ; APPLICANT: Liu, Chenghua
 ; APPLICANT: Asundi, Vinod
 ; APPLICANT: Zhang, Jie
 ; APPLICANT: Drmanac, Radotje T.
 ; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
 ; FILE OF INVENTION: Polypeptides
 ; FILE REFERENCE: 784CIP2BCIP
 ; CURRENT APPLICATION NUMBER: US/10/117,722
 ; PRIORITY FILING DATE: 2002-04-04
 ; PRIOR APPLICATION NUMBER: 09/620,312
 ; PRIOR FILING DATE: 2000-07-19
 ; PRIOR APPLICATION NUMBER: 09/552,317
 ; PRIOR FILING DATE: 2000-04-25
 ; PRIOR APPLICATION NUMBER: 09/488,725
 ; PRIOR FILING DATE: 2000-01-21
 ; NUMBER OF SEQ ID NOS: 1104
 ; SOFTWARE: pt_FL_genes Version 1.0
 ; SEQ ID NO 155
 ; LENGTH: 6990
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (61)..(2397)
 US-10-117-722-155

Alignment Scores:

Pred. No.: 1.09e-32 Length: 6990
 Score: 571.50 Matches: 196
 Percent Similarity: 41.68% Conservative: 137
 Best Local Similarity: 24.53% Mismatches: 305
 Query Match: 10.84% Indels: 161
 DB: 17 Gaps: 25

US-09-914-042-1 (1-1006) x US-10-117-722-155 (1-6990)

QY 38 AlaAlaIleGluGluAlaLeu---AspValAspArgMetValLeuTyrLysMetLysLys 56
 Db 115 GCAGCTTTGGAGAAAGTAGAGGTGATGTGGCAGAAATTTGGAACTA---AAACTTGATAAG 171
 QY 57 SerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThr 76
 Db 172 CTTGTGAACCTT-----GGGATTGCAATGATTGATCTGCGAAAGCCCTTTGT 219
 QY 77 GlnAlaLeuGluLysPheGlyGlyAsnCysValCysArgAsp----- 90
 Db 220 GTTGCAATAAACAAGTTCATGAATGGG-----ATTGAGACCTGGCCCGCATTTCTAGT 273
 QY 91 ---AspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
 Db 274 AATGATGCTGCTGTGAGACAAAGTTTGACCAAGTTTTCTGACAGCTTTCACAGAAATGATA 333
 QY 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnIleIleSerPheProLeuAspSer 129
 Db 334 AATTTTCACACATCTGTTGCCAACTCAGAGATCAATTTAAGCCACAGCTTCAGTAAC 393
 QY 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysPheAspLysAla 149
 Db 394 TTTGTTAAAGAGATCTTAGAAATTCAAA---GATGCCAAGAACAATTTGAAAGATC 450
 QY 150 TrpLysAspTyrGluThrLysIleThrLys-----IleGluLysGluLysGlu 166
 Db 451 AGTGAAGAAAGAAATTCGCTTAGTAAATAATGCCAAGTACCAAGAAACAAACACAT 510

QY 167 HisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGlu 186
 Db 511 GAAGTTGAA-----GAAGCCACCAACAT 534
 QY 187 MetGluLysGluArgArgPheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsn 206
 Db 535 CTGACAGCAACAGAAATGTTCCGACACATAGCCCTCGATATATGTGCTTCAGATTAAT 594
 QY 207 GluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAla 226
 Db 595 GTTCTTCAATCAAAAAGGAGATCAGAAATCTTAAATCAATGTTGCTTATGATGCC 654
 QY 227 GlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGlu 246
 Db 655 CATTGGCCTTCTTTTCATCAAGGATATGATCTGTTAGTGAACCTTGGACCTTACATGAAG 714
 QY 247 ThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGlnLeu 266
 Db 715 GATCTTGGTCACAGTTGGATCGACTGTTGATGCGAGCAAGAGAGAAAGAAATG 774
 QY 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlnLysGluAspSerGln 286
 Db 775 GAGCAAAAACATTC-----ACCATTCAACAAAAGATTTCTCCAGT 816
 QY 287 IleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306
 Db 817 GATGATTTCTAAGTTAGAAATATAAGCTA-----GATGCTGCAATGGCATA 861
 QY 307 GluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArg 326
 Db 862 GTTATGGAAGGATATCTGTTCAACAGCCAGCAATGCTTCAAACTTTGGAACAGGCGC 921
 QY 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346
 Db 922 TGGTTTTCAATACAGATAATCAGTTGTTTACCAGAAAAAATTTAAGATATCCGACT 981
 QY 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluLys 363
 Db 982 GTGGTAGTTGAAGACCTCAGGCTTTGCACAGTGAACATTTGTGAAGACATAGAGCCACGA 1041
 QY 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGln 383
 Db 1042 TTCTGCTTTGAGGTGCTCGCCCAACAAAAAGTTGTCATGCTCCAGCGAGATTTCCGAAAAG 1101
 QY 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403
 Db 1102 CTGCGCCAGGCATGATTAAGCTGTTCCAGACCATTTGCTACTGCTTTATAGAGAG--- 1158
 QY 404 PheLysGlyAspAsp-----AsnThrGlyGlu--- 412
 Db 1159 ---AAGGCTGATGAATCAGAGAAGCTGGATAAGAAATCATCTCCATCCACAGGAAGCCTA 1215
 QY 413 -----AsnAsnIleValGlnLeuThrLys-----GluIleIleSerGluVal 427
 Db 1216 GATTCTGGAATAGTCCAAAGAGAAATTTTGAAGAGGAGAAAGTGCCTTTCCAGCGGTC 1275
 QY 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447
 Db 1276 CAGTGATTCCTCGCAATGCCAGCTGTTGTGACTGTGSCCTGGCAGATCCACGGTGGGCC 1335
 QY 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467
 Db 1336 AGCATCAACCTGGGCATCACCTTGTGATCGAGTCTCCGGAATTCACGGAGCCCTTGGG 1395
 QY 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487
 Db 1396 GTTCATTTTCAAAGTACGATCTTAACTTTAGACACCTGGGAGCCAGAACTTTTAAAG 1455
 QY 488 LeuAlaLysAsnIleGlyAspAlaGlyPheAsnGluIleMetGluCysLeuProAla 507
 Db 1456 CTTATGTGTGAGTTGGGGAATGATGTTATATAAATCGAGTTTATGAAGCTTAATGTGGAANA 1515

Qy 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527
 Db 1516 ATGGGATAAAGAACCCCAACGAGG---CAAGACAGGAGAGGAGGCAATATATACGA 1572
 Qy 528 AlalysTyrIleGluArgArgTyrAla 536
 Db 1573 GCAAAATATGTGGAGAGGAAATTTGTGGATAAATATTCTATATCATTTATCCTCCTGAG 1632
 Qy 537 ---ArgLysLysHisAlaAspAsnAlaAlalysLeuHisSerLeuCys----- 551
 Db 1633 CAGCAAAAAAAGTTGTCTCTAAAAGTTCTGAAGAAAAAGAGCTGAGCATTTCTAAATTT 1692
 Qy 552 -----GluAlaValLysThrArgAspIlePhe 560
 Db 1693 GGGCCAGGGACCAAGTCAGAGCATCTCCCAAGTTTCAGTCAGAGGAAATGAC---AGT 1749
 Qy 561 GlyLeuGluGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 1750 GGAATTCAGCAGAGCTCTGTATGATGGAAGAGAACTCTTACCTCCACGGTGTGACCAAT 1809
 Qy 581 Gly---HisGluProAsp----- 585
 Db 1810 AGTTTATAGCCTCTGAAGGAGAAAGCAAGATTCTTCTATGTTTCTTGACTCGAAACAT 1869
 Qy 585 ----- 585
 Db 1870 CTTAATCAGACATTCAGCTTATAGGGCGTCATATGAAAAAACCTTCCTAAATGGCT 1929
 Qy 586 GluThrAlaLeuHisLeuAla----- 592
 Db 1930 GAGGCTTTGGCTCATGGTGCAGACGTGAACCTGGGCAATTTCAGAGAAAAACAAGCGACA 1989
 Qy 593 -----ValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsn 610
 Db 1990 CCACCTATTACGGCTGTATTAGGGGCTCTTTGGTGACGTGTGAGTTCTCTCTACAGAAAT 2049
 Qy 611 SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeu 630
 Db 2050 GGTGTAATGTCAACAAAGAGATGTCGAAGGCGGGGACCATTCACCATGCCACCGCTC 2109
 Qy 631 ThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAla 650
 Db 2110 TTAGGCGACACAGGCGAGGTATGTTTATCTTAAACGAGGTGCCCAATCAACATGCCACT 2169
 Qy 651 AsnGluSerGlyLysThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlu 670
 Db 2170 GATGAAGAAGGGAAGACCCCTTTGACATAGCTGTGGAAGCAGGCAATGCTGATATAGTC 2229
 Qy 671 GluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGlu 690
 Db 2230 ACCTTGTTACGTTTAGCA----- 2247
 Qy 691 TrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
 Db 2248 -----AGAATGAATGAAGAGATCGGGAATCAGAAGGACTTTATGGA-----CAG 2292
 Qy 711 ProSerGluAsnArgArgGluAspArgProLeuSerPheTyrGlnLeuGlySerAsn 729
 Db 2293 CCAGGTGATGAACCTTATCAGGACATATTTCTGTGATTTTTCCTCCAAATGGCATCCAAT 2349

RESULT 13

US-10-080-334-79
 ; Sequence 79, Application US/10080334
 ; Publication No. US20040002584A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pena, Carol E. A.
 ; APPLICANT: Shimkets, Richard A.
 ; APPLICANT: Li, Li
 ; APPLICANT: Shenoy, Suresh G
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Spytek, Kimberly A.
 ; APPLICANT: Vernet, Corine A. M.
 ; APPLICANT: Malyankar, Uriel M

; APPLICANT: Guo, Xiaojia
 ; APPLICANT: Gusev, Vladimir Y
 ; APPLICANT: Casman, Stacie J
 ; APPLICANT: Boldog, Ferenc L
 ; APPLICANT: Furtak, Katarzyna
 ; APPLICANT: Tchernev, Velizar T
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Gangolli, Esha A
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Liu, Xiaohong
 ; APPLICANT: Baumgartner, Jason C.
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Spaderna, Steven K
 ; APPLICANT: Zerhusen, Bryan D
 ; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
 ; FILE REFERENCE: 21402-275
 ; CURRENT APPLICATION NUMBER: US/10/080,334
 ; PRIOR FILING DATE: 2002-02-21
 ; PRIOR APPLICATION NUMBER: 60/270,523
 ; PRIOR FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: 60/322,712
 ; PRIOR FILING DATE: 2001-09-17
 ; PRIOR APPLICATION NUMBER: 60/311,980
 ; PRIOR FILING DATE: 2001-08-13
 ; PRIOR APPLICATION NUMBER: 60/330,307
 ; PRIOR FILING DATE: 2001-10-18
 ; PRIOR APPLICATION NUMBER: 60/278,796
 ; PRIOR FILING DATE: 2001-03-26
 ; PRIOR APPLICATION NUMBER: 60/281,521
 ; PRIOR FILING DATE: 2001-04-04
 ; PRIOR APPLICATION NUMBER: 60/276,677
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/311,595
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: 60/270,220
 ; PRIOR FILING DATE: 2001-02-21
 ; PRIOR APPLICATION NUMBER: 60/274,295
 ; PRIOR FILING DATE: 2001-03-08
 ; PRIOR APPLICATION NUMBER: 60/318,526
 ; PRIOR FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: 60/286,548
 ; PRIOR FILING DATE: 2001-04-25
 ; PRIOR APPLICATION NUMBER: 60/291,765
 ; PRIOR FILING DATE: 2001-05-17
 ; PRIOR APPLICATION NUMBER: 60/270,797
 ; PRIOR FILING DATE: 2001-02-23
 ; PRIOR APPLICATION NUMBER: 60/276,400
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/270,810
 ; PRIOR FILING DATE: 2001-02-23
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 79
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-080-334-79

Alignment Scores:

Pred. No.:	3,48e-33	Length:	2484
Score:	570.50	Matches:	212
Percent Similarity:	36.49%	Conservative:	143
Best Local Similarity:	21.79%	Mismatches:	364
Query Match:	10.82%	Indels:	254
DB:	17	Gaps:	24

US-09-914-042-1 (1-1006) x US-10-080-334-79 (1-2484)

Qy 23 AlaSerSerPheThrThrArgThrAlaGlnCys-----ArgAsn 35

Db 6 GCCGCGCCATGACCGTGGAGTTCGAGAAAGTCCGTCAGAACTCCCGCGCTTCAGGGCG 65


```
Db 2118 CCGTTGGCCATCGCA-----GTGCGAGCG 2141
Qy 677 LeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGlu 696
Db 2142 GCCAAC----- 2147
Qy 697 AspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArg 716
Db 2147 ----- 2147
Qy 717 GluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaVal 736
Db 2148 -----GCTGACATCGTG 2159
Qy 737 SerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSer 756
Db 2160 ACACGTGCTCCGT--CTGCGCGCATGCGGAGAAATGCGCGAG----- 2201
Qy 757 IleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnPro 776
Db 2202 -----GCCAGGCTGCCCTCGTCTCCCGCGCGCGCTGGCG 2237
Qy 777 -AlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValG1 796
Db 2238 GGCAGCCCCACGAGCTCCAGTTCC-----GCAGGTGTATC 2273
Qy 796 nThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlyse 816
Db 2274 CAGGAGTTCATCAGCTCCACC--TGGAAAG----- 2301
Qy 816 rArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLeuArgVa 836
Db 2302 -----AGAGTAGGCGCGGCGGCGGCGGAGCTGCACCCCGCGCGCGCGAGCG-- 2352
Qy 836 lThrSerThrAsnProLeuThrProThrProProProValAlaLysThrProSerVa 856
Db 2353 -----CCGCGATGCCCGGAGTCCCTGGCGCGCGCGCGCGCGCG 2390
Qy 856 lMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLleSerGlnIleArgPr 876
Db 2391 GGCCTGCTGTGTCGCGCGGTCTGATCTGCGAGCGCC-----AGTGTGGGCGCC 2444
Qy 876 oProProLeuProProGlnProProSerArgLeuPro 888
Db 2445 GCGGCCCTGTGTCGCCAGGAGAGGAGCGAGGCCGCCCA 2481

RESULT 14
US-10-176-306-10
; Sequence 10, Application US/10176306
; Publication No US20030130485A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Curtis, Rory A. J.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Kapeller-libermann, Rosana
; TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
; FILE REFERENCE: 10448-195001
; CURRENT APPLICATION NUMBER: US/10/176,306
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/001,137
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: PCT/US01/45291
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/248,362
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,331
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/248,365
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: 60/250,077
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,327
```

```
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 60/250,176
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 10/023,617
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/49416
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/256,249
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,405
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 10/083,248
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: PCT/US01/46717
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/242,324
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/242,518
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/241,989
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 2737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (69)...(2570)
US-10-176-306-10

Alignment Scores:
Pred. No.: 2,07e-32 Length: 2737
Score: 561.00 Matches: 232
Percent Similarity: 37.13% Conservative: 136
Best Local Similarity: 23.41% Mismatches: 425
Query Match: 10.64% Indels: 199
DB: 15 Gaps: 32

US-09-914-042-1 (1-1006) x US-10-176-306-10 (1-2737)
Qy 20 AlaProThrAlaSerSerPheThrThrArgThrAlaGlnCys----- 33
Db 51 GCGCGCGCGCGCGCGCGCATGACCGTGGAGTTCAGAGAGTGGTCAAGGACTCCCGCGC 110
Qy 34 ---ArgAsnThrValAlaAlaIleGluGluAlaLeuAspValAspArgMetValLeuTyr 52
Db 111 TTCAGGCGGACCATTCAGAGGTGGAGACGCGAGATTGGAGATTGAGGCCAAACTGGAC 170
Qy 53 LysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGlu 72
Db 171 AAGCTGTGAAGCTGTGCAGTGGCATGTGTGAAGCCGCTAAGCGCTACGTGAGCCAGC 230
Qy 73 GluGlnTyrThrGlnAlaLeuGluLysPheGlyAsnCysValCysArgAspPro 92
Db 231 AGGCTTTTCGTGACGCGCGCTCCGACCTGCCAGAG-----TGCAGGGCGACACC 284
Qy 93 AspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThrAlaLeuPhe 112
Db 285 GTCATCTCGGAATGCTTCAGAGGTTCCTGCACAGCCTACAGGAGGTGGTGAACCTACCAC 344
Qy 113 LysAsnLeulleGlnAsnMetAsnIleIleSerPheProLeuAspSerLeuLys 132
Db 345 ATGATCTCTGTTTGACAGGCTCCGTCGCGGACGAGCTCCAGAGCTTTGTCAAA 404
Qy 133 GlyAspLeuLysGlyValLysGlyAspLeuLysProPheAspLysAlaTrpLysAsp 152
Db 405 GAGGATGTCGGAAGCTTCAG--GAGACAAGAGGAGGTTTTCAGAGTTCGGGAGGAC 461
Qy 153 TyrGluThrLysIleThrLysIleGluLysLysGluHisAlaLysLeuHisGly 172
Db 462 CTGAGGCTGTCCCTGGTG-----AGGAACGCCCGCGCGCGCGCGCGCGCGG 506
```


QY 173 MetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLysGluArgArg 192
DB 507 CCCACGAGGTGGAG-----GAAGCACCAGCGGGCCCTCACCTCCACGAGGAAG 554
QY 193 PhePheGlnLeuGlnMetCysGluIleValLeuLeuLysValAsnGluIleLysIleLysLys 212
DB 555 TGCTTCCCGCCACCTGGCAGTGTGCTGCTCCAGATCAATGTTCTGCAGGCCAAGAAG 614
QY 213 GlyValAspLeuLeuGlnAsnLeuIleLysIlePheHisAlaGlnCysAsnPhePheGln 232
DB 615 AAGTTTGAGATCTCGACTCTATGCTGTCTTCATGCGCAGCCAGCTCCAGCTTCTCCAG 674
QY 233 AspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeu 252
DB 675 CAGGCTACAGCTCTCTGACCGAGTGGACCTCTACATGAAGAAGTGGCAGCCGAGCTG 734
QY 253 -----HisThrIleLysGlnAlaGlnAspGluGluArgGlnLeu 266
DB 735 GACCAGCTGGTATCGACTCTCGGTGGAAAGCTGTGATGGAGCGGAAAGCAGCCGCC 794
QY 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGln 286
DB 795 ATCCAGCAGCGGACCTGCTGAGGACTTCTCTACGATGAGTCCAAAGTGGAGTTTAC 854
QY 287 IleArgGlnSerThrAlaTySerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306
DB 855 GTG-----GACGCGCCAGTGGGGTG 875
QY 307 GluArgAsnGlySerLeuTyIleLysSerAspGlyIleArgLysValTrpGlnLysArg 326
DB 876 GTGATGGAGGGTACTCTTCAAGAGGGCCAGCAACGCTTTCAAGACATGGAACCGCGC 935
QY 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346
DB 936 TGGTTCTCATTCAGAACGAGCAGTGTGTACCAAGAAAGCTCAAGATGCCCTCAC 995
QY 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluLys 363
DB 996 GTGGTGGTGGATGACCTCGCGTGTCTGTGAAGCGGTGTGAGGACATCGAGCGGAG 1055
QY 364 LysCysPheAspLeuIleSerHisAspArgThrTyHisPheGlnAlaGluAspGln 383
DB 1056 TTCTGCTTTCGAGTGTCTGCTCCCAAGAGAGTGCATGCTGCAGGCTGACTCCGAGA 1115
QY 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu----- 400
DB 1116 CTGCGGCAAGCTGGGTCCAGGCTGTGAGGCGCAGCATCGCTCCGCTACCGCGAGAGC 1175
QY 401 -----AsnAsnAlaPhe 404
DB 1176 CTGACAGTTGCTATAGCAGAGGCTGGACCGCACAGCATCCCGTCCACGAGCAGCATC 1235
QY 405 LysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIle 424
DB 1236 GACTCCGCCACCGACACTCGGAGCGGTGGCGTGAAGGGCGAG-----AGTGTGCTG 1286
QY 425 SerGluValGlnArgMetThrGlyAsnAspValCysAspCysGlyValAlaProAspPro 444
DB 1287 CAGCGTGTGCAGATGTGCCCGGCAACAGCCAGTGGCGGAGTGGCGGAGCGCGGACCCC 1346
QY 445 ThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArg 464
DB 1347 CGCTGGCGCAGCATCAACTGGCGGTGTGCTCTGCTGATGAGTCTCCGGCATCCACAG 1406
QY 465 GluLeuGlyValHisTySerProMetGlnSerLeuThrLeuAspValLeuGlyThrSer 484
DB 1407 AGCCTGGGTGTCCACTGTCTCAAGGTGGGTCCCTGACGCTGAGCTCGTGGAGGCTGAG 1466
QY 485 GluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCys 504
DB 1467 CTGTAAAGCTGTGTGAGCTTGGAAACAGCGCTGTGAATCAGATCTATGAGGCCAG 1526

QY 505 LeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAsp 524
DB 1527 TGTGAGGTGCGAGCAGCAGGAAACCCACAGCAGCAGCTCCCGCAGCAGCAAGAGGCC 1586
QY 525 TyrIleThrAlaIleValTyIleGluArgArgTyAlaArgLysValHisAlaAspAsnAla 544
DB 1587 TGGATCAAGGACAATACGTGTGAAGAAGTTTCTGCGGAAG----- 1628
QY 545 AlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePheGlyLeuGln 564
DB 1629 GCGCCATGGCAGCCCTCGAGGCCCCAAGAGCGTGGAGGGTGCAGAGTGCCTGCGG 1688
QY 565 AlaTyAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluPro 584
DB 1689 CCCACAGC-----TCTCCCGCGCTCCCACTGCCCGCGCAAGGTCCGG 1733
QY 595 AspGluThrAlaLeuHis-----LeuAlaValArgSerValAspArgThrSerLeuHis 602
DB 1734 CTTGAGCCCGTTCCTGCTGTGCGCTCTGTCTCAGTG----- 1775
QY 603 IleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySer 622
DB 1776 -----GGCACCTGGATCGTAAGTTCCGCGCAGACTCC 1808
QY 623 ThrAlaLeuHisTyCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuArg 642
DB 1809 CTC-----TTCGT-----CCGACGAGCTGGACTCGCTC----- 1838
QY 643 GlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLys 662
DB 1839 -----TTCCTCTACTCGACGAGGGCCCGAGGGCTGGCCCTCGCAGTCTGTAGTAGC 1892
QY 663 ArgLeuLysHisGluHisCysGluGluLeuThrGlnAlaLeuSerGlyArgPheAsn 682
DB 1893 -----GACAGTGGCCTTGGGGCAGCTCGGAT 1919
QY 683 SerHisValHisVal-----GluTyr 689
DB 1920 GGCAGCTCGGACGCTCTGGCTTCGGCTCGGGCTCTGTGGTGGACAGCGTCACTGAGGAG 1979
QY 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
DB 1980 GAGGTCGAGAGTCCGAGGAGTCCAGCGGTGAGGACAGCGGGACACTGAGGCCGAGGCC 2039
QY 710 GlnProSerGluAsnArgArgGluAspArgPro-----IleSerPheTyTrpGlnLeuGly 727
DB 2040 TGGGGCTTGGCGGACGCTCGGAGCTGCACCCGGGGCTCTTGGCGCACCGCGCGCGCT 2099
QY 728 SerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAla-----AlaAsnLeuAla 746
DB 2100 GCGCGGACCTCTCTGGCGCGCTGGCGCGCTGGCCACCGGGCGAGGTCAACTGGGCG 2159
QY 747 LysGluLysGlnArgAlaPheMetPro-----SerIleLeu 758
DB 2160 GACCGGAGGATGAGGCAAGACCGCTGTGTGCGGCGCTGTGAGGGGCTCTGTATC 2219
QY 759 GlnAsnGluThrTyTrpGlyAlaLeuSerGlySerProProAlaGlnProAla--- 777
DB 2220 GTCTGTGATTCCT-GCTGCAAAAGCGAGCGGAGCTGAGAACCAAGAGACACCGCGGGCG 2278
QY 778 -----AlaProSerThrThrSerAlaPro-----ProLeuPro--- 788
DB 2279 GCGCGCTGCGACCGCAGCTGTGGCGCGCACCGCGGAGTTTGCCTGTTCTGAA 2338
QY 789 -----ProArgAsnValGlyLysValGlnThrAlaSer 799
DB 2339 GCGGGCGCGACCGACGCGCTGAGCAAGACAGCGGAGCCCGTTGGCCATCGCAGT 2398
QY 800 SerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArg 819
DB 2399 GCA-----GGCGGCAACGCTGACAT 2419
QY 820 SerSerAspProProAlaValHisPro-----Pro 830

```

Db 2420 CGTGACACTGCTCCGCTGCGCGCATGCGGAGGAATGCGGAGGCGGAGGCTGCCCC 2479
Qy 831 LeuProProLeuArgValThrSerThrAsnProLeuThrProThrProProProVal 850
Db 2480 TGGTCCCGCGGCGC-----CCTGGCGGCGAGCCGCCACCGA 2515
Qy 851 AlalysThrProSerValMetGluAlaLeuSerGlnProSerLys----- 865
Db 2516 GCTCAGTTCGCGAGGTATCCAGGAGTTATCAGCCTCCACCTGGAGAGAGCTAGGG 2575
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProProLeuProGlnProProSer 885
Db 2576 CCGGCGCGCGCGGCGAGTGCACCCCGCGCGCGCGCGCGCGCGCGCGAGTCC 2635
Qy 886 ArgLeuProGlnLysLysProAlaPro-----GlyThrAspLysSerThr 900
Db 2636 CTGGCGCC--ACCGCGCGCGCGCTCGCTGTGACCGCGGCTCGATACCTGCGCAGCC 2692
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyPro 911
Db 2693 CCAGTGTGGGCGCGCGCGCTCTGCGCCA 2725

RESULT 15
US-10-723-860-8033
; Sequence 8033, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8033
; LENGTH: 5421
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1861)..(1865)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2487)..(2509)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-8033

Alignment Scores:
Pred. No.: 8 3e-32 Length: 5421
Score: 558.00 Matches: 256
Percent Similarity: 34.16% Conservative: 156
Best Local Similarity: 21.23% Mismatches: 422
Query Match: 10.58% Indels: 374
DB: 20 Gaps: 38

US-09-914-042-1 (1-1006) x US-10-723-860-8033 (1-5421)
Qy 12 AlaGluThrHisGluAspTyrLysAlaProThrAlaSerSerPheThrArgThrAla 31
Db 1743 GCTGAGGCTGACGGTGACCCAGCCAGCC---CGTTCTTGTGTTAGGCGCAGCATTCAC 1799
Qy 32 GlnCysArgAsnThrValAlaAlaIleGluAlaLeuAspValAspMetValLeu 51
Db 1800 GAGGTGGAGACGGACGTGGTGAGATTGAGGCCAAACTGGAC----- 1841
Qy 52 TyrLysMetLysLysSerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsn 71

```

```

Db 1842 ---AAGTGGTGAAGCTGTGCANNNNATGGTGAAGCCGGTAAGGCTACGTGAGCACC 1898
Qy 72 GluGluGlnTyrThrGlnAlaLeuGluLysPheGlyGlyAsnCysValCysArgAspAsp 91
Db 1899 AGCAGGCTTTTGTGTGAGCGGCGCTCCGACACCTGTCCAGCAGAG-----TGCAGGCGGAC 1952
Qy 92 ProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThrAlaLeu 111
Db 1953 ACCGTATCTTCGGAATGTCTGCAGAGGTTCCTGCAGAGCTTACAGGAGGTGGTGAACCTAC 2012
Qy 112 PheLysAsnLeuIleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSerLeuLeu 131
Db 2013 CACATGATCTCTGTTTACCAGGCCAGAGGTCCGTGCGGCAGCAGCTCCAGAGCTTGTCTC 2072
Qy 132 LysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPheAspLysAlaTyrLys 151
Db 2073 AAAGAGGATGTCCGGAAGTTCAAG--GAGACAAGAAGCAGATTGTGACAGGTGCGGAG 2129
Qy 152 AspTyrGluThrLysIleThrLysIleGluLysGluLysLysGluHisAlaLysLeuHis 171
Db 2130 GACCTGGAGCTGCTCCCTGGTG-----AGGAACGCCAGGCCCGCCAGGCGAC 2174
Qy 172 GlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGluGluMetGluLysGluArg 191
Db 2175 CGGCCCCAGCAGGTGGAG-----GAAGCCACCGGGGCCCTCACCTCCACGAG 2222
Qy 192 ArgPhePheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIleLys 211
Db 2223 AAGTGTTCGCCACCTCGGCACCTGATGTCTCCAGATCAATGTCTTCGAGGCCAAG 2282
Qy 212 LysGlyValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPhePhe 231
Db 2283 AAGAAGTTTGAGTCTCGGACTCTATGCTGCTTCATGCACGCCCGCTCCAGCTTCTTC 2342
Qy 232 GlnAspGlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAsp 251
Db 2343 CAGCAGGCTACAGCTCTCTGCACAGCTGGACCCCTACATGAAGAAGCTGGCAGCCGAG 2402
Qy 252 Leu-----HisThrIleLysGlnAlaGlnAspGluGluArgArgGln 265
Db 2403 CTGACACCAAGCTGTGTGATCGACTCTGCGGTGGAAGCGTGAGATGGAGCAAGCAGGCC 2462
Qy 266 LeuIleGlnLeuArgAspIleLeu-LysSerAlaLeuGlnValGluGlnLysGluAspSe 285
Db 2463 GCCATCCAGCAGCGGACGCTGCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2522
Qy 285 rGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGln 305
Db 2523 TGACGTG-----GAGCGCGCCAGTGG 2543
Qy 305 YThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArgLysValTyrGlnLys 325
Db 2544 GGTGTGATGAGGCGCTACCTCTTCAAGAGGCGCCAGCAACGCTTTTCAAGACATGGAACCC 2603
Qy 325 sArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgPr 345
Db 2604 GCGCTGTTCTCCATTGAGAACAGCCAGCTGGTCTACAGAGAAGCTCAAGGATGGCCT 2663
Qy 345 oProAlaLysLeu--AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGln 362
Db 2664 CACCGTGTGTGGATGATCCCTCCGCTGTCTGTGAAGCCGCTGAGGACATCGAGCG 2723
Qy 362 uLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGln 382
Db 2724 GAGGTTCTGCTTCAGGTGTGTCCACCAAGAGCTGATGTGTGAGGCTGATCCGGA 2783
Qy 382 uGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu----- 400
Db 2784 GAAGCTGGGCAACCTGGGTCCAGGCTGTGACAGGCCAGCATCCCTCCGCTTACCGCGA 2843
Qy 401 -----AsnAsnAla 403

```

Db 2844 GAGCCTGACAGTTGCTATAGCGAGAGGCTGACCGCACAGCATCCCGTCCACGAGCAG 2903
 Qy 403 aPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluI 423
 Db 2904 CATCGACTCCGCGCACCGACACTCGGAGCGGTGGCGTGAAGGCGGAG-----AGTGT 2954
 Qy 423 eIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAs 443
 Db 2955 GCTGACGCTGTGCAGAGTGTGGCGGCAACAGCAGTGCAGCGGCGGCGGCGGCGGCGG 3014
 Qy 443 pProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHi 463
 Db 3015 CCCCCTGGGCGCAGCATCACTGGGCGGTGCTCTGCTGATGAGTGTCTGGGATCA 3074
 Qy 463 sArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyTh 483
 Db 3075 CAGGAGCCTGGGTGCTCCACTGCTCCAGAGTGGGTCCCTGAGCGTGGAGTCTGTGGGAGCC 3134
 Qy 483 rSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCy 503
 Db 3135 TGAGCTGTAAAGCTGATGTGTGAGCTTGGAAACAGCGTGTGAATCAGATCTATGAGGC 3194
 Qy 503 sCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLy 523
 Db 3195 CCAGTGTGAGGTGCGAGCAGCAGGAAACCCACAGCAGCAGTCCCGGCGAGCAAGGA 3254
 Qy 523 sAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLys----- 538
 Db 3255 GGCCTGGATCAAGCAATAGTGTGGAAGAGTTTCTGCGAAGGCGGCCATGCGACC 3314
 Qy 538 ----- 538
 Db 3315 AGCCTGGAGGCCCCAAGACGCTGGAGGTGAGAAAGTGCCTGCGGCCCCACAGCTCTCC 3374
 Qy 538 ----- 538
 Db 3375 CCGCGCTCCCCTGCGCGGCAAGTCCGGCTTGAGCCGCTTGCCTGTGTGGCGCGC 3434
 Qy 538 ----- 538
 Db 3435 TCTGTCTCAGTGGGACCTGGATCGTAAGTTCCGCGCAGACTCCCTCTTCTGTCCCGA 3494
 Qy 538 ----- 538
 Db 3495 CGAGCTGAGACTGCTTCTCTACTTCCAGCGAGGGCGCGAGGGGTGGCCCTCGCAA 3554
 Qy 539 -----LysHisAlaAspAsnAlaAlaLysLeuHi 548
 Db 3555 GGGTGCAGAGTGGAGGAGTCCAGCGGTGAGGACAGCGGGGACACTGAGGCGCGGCGCTG 3614
 Qy 548 sSerLeuCysGlu-----AlaValLysTh 556
 Db 3615 GGGCTCGGAGCTGCGGAGCTCCAGCCGGGGCTTGTGGGCGACCGCGCAGCGCGTGC 3674
 Qy 556 rArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysI 576
 Db 3675 CCGCGACCTTCTGCTGCTGGGCGGCGTGCSCCCACGCGGCGGAG-----GT 3722
 Qy 576 eProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerVa 596
 Db 3723 CAACTGGGCGGAGCGGAGGATGAGGCAAGACGCGGCTG-----GTGACGCGCT 3773
 Qy 596 lAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLy 616
 Db 3774 GCTAGGGGCTCTCTTGATGCTGTGTGAGTTCTGTGCAAAACGAGCGGAGCGTGAACCA 3833
 Qy 616 sGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCy 636
 Db 3834 AAGACACAGCGGCGCGGCGCCCTGTCACACCGCCAGCTGTGTGGCGCGCACCGGCCA 3893
 Qy 636 sLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluTh 656
 Db 3894 GCTTTGCTGTTCTGAAAGCGGGGCGGAGCCAGCACGCGCTTGGACCAAGAGCAGCGGGA 3953

Qy 656 rProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAl 676
 Db 3954 CCGTGTGCCATCGCA-----GTGACGCG 3977
 Qy 676 aLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGl 696
 Db 3978 GGCCAA----- 3984
 Qy 696 uAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgAr 716
 Db 3984 ----- 3984
 Qy 716 gGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaVa 736
 Db 3985 -----GCTGACATCGT 3995
 Qy 736 lSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSe 756
 Db 3996 GACACTGCTCGT---CTGGCGCGCATGGCGGAGGAATGCGCGAG----- 4038
 Qy 756 rIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnPr 776
 Db 4039 -----GCCAGGCTGCCCTCGTCCCGGGCGGCGCTCGC 4073
 Qy 776 o-AlaAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLysValG 796
 Db 4074 GGGCAGCCCCACGAGCTCCAGTTCC-----GCAGGTGTAT 4109
 Qy 796 lNThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGly---- 814
 Db 4110 CCAGGATTATCATAGCTCCACC---TGGAAAGAGAGCTAGGGCGGCGGCGGCGGAGC 4166
 Qy 815 -----GlySerArgGlnArgSerSerSerSerSerSerProAlaValHisProL 831
 Db 4167 TGCACCCCG 4223
 Qy 831 euProProLeuArgValThr-----SerThrA 840
 Db 4224 CCGCGCGCGCTGGGTGTGACCGCGGGTCGATACCTGGCGAGCCCCAGTGTGGGGCGCGC 4283
 Qy 840 snProLeuThrPro----- 844
 Db 4284 GGCCTGTGCTGCCCGGAGGAGGAGAGCGGGCGCCACACTAGTGTCTTTGAAGCCTCAGCT 4343
 Qy 845 -----ThrProProPro-----ProValA 851
 Db 4344 TTCCCTGGGGGGGTGCTGCTGCTGCGGTGTCCCTCACCCACCTGGGGGAACCTCTGTCT 4403
 Qy 851 laLysThrPro-----SerValMetGluAlaLeuSerGlnProS 864
 Db 4404 TCAGGTCACTCCCTTTTCAGGGGCTGGGTGCTCATGTACAGCCACTTTTAAAGCCCG 4463
 Qy 864 erLysPro-----AlaProGlyIleSerGlnIleArgPro----- 876
 Db 4464 TGTCACTGTCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 4523
 Qy 877 -----ProProLeu-ProProGlnProProSerArgLeuPro 888
 Db 4524 AACACTGACCTGTAGTGGTGTCTCTCTGTGACACCCCGAGAGCTCTCTCCAGG---GCA 4580
 Qy 889 GlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnPro 908
 Db 4581 GAAAGGTGGCCCATCAGGGAACCAAGAGTACTCTACTGTAGTACC----- 4628
 Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetVal 928
 Db 4629 -----CCAGCATCCAGCTGCGGTGCTGGCGGAGGGGCTCTGGCTGGGT 4682
 Qy 929 LeuGlnPro-----ProAlaProMetProArg---LysSerGlnAlaThrLysLeuLys 945
 Db 4683 CCTCAGCGCTGGTGGAGCGCGGGGTGCCAAGGCCAGCTGGGTG---GCCACGAAGTCCCC 4741

Qy	946	ProLysArgValLysAlaLeuTyrAsnCysValalalaAspAsnProAspGluLeuThrPhe	965
	4742	-----GTGCACAGACCCCGCCATCTCCGGCC	4771
Db		:::	
Qy	966	SerGluGlyAspValIlelleValaspGlyClucluaspGlnGluTrpTrlleGlyHis	985
Db	4772	TCTTCGGAGGCACCTTC-----TCCTGGTA-CTCGGCCCT	4806
Qy	986	IleaspGlyaspProGlyArgLyasglyAlaPheProValSerPheValHis-----	1002
Db	4807	GTG-----CCTGGCATAAAGCCCCTTTGGCACTTTGCAGCCCTCATCTCATGTTCT	4857
Qy	1003	PheIleAlaasp	1006
Db	4858	TTTGATCGGCAC	4869

Search completed: August 4, 2005, 17:14:51
Job time : 2828 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 08:33:01 ; Search time 393 Seconds
(without alignments)
4188.535 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGFVSVFHFAD 1006

Scoring table:

	BLOSUM62	Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5		
Fgapop 6.0 , Fgapext 7.0		
Delop 6.0 , Delext 7.0		

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US09914042/runat_04082005_090633_2136/app_query.fasta_1.1159
-DB=Issued Patents NA -QWTF=fastcap -SUFFIX=rni -MINWATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09914042_@CGN 1 1 80 @runat_04082005_090633_2136 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6C.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3874	73.5	2949	4	US-09-023-905A-8
2	3874	73.5	5954	4	US-09-023-905A-6
3	3691	70.0	4486	1	US-08-322-742-15
4	3417	64.8	4328	1	US-08-322-742-12
5	3155	59.8	2712	4	US-09-023-905A-11
6	3155	59.8	4595	4	US-09-023-905A-9
7	3013.5	57.1	5330	4	US-09-023-905A-1
8	2954.5	56.0	3456	4	US-09-023-905A-5
9	2954.5	56.0	4382	4	US-09-023-905A-3
10	571.5	10.8	6990	4	US-09-620-312D-155
c 11	453	8.6	1090	4	US-09-270-767-1629
c 12	453	8.6	1090	4	US-09-270-767-16911

c 13	429	8.1	2041	4	US-09-828-303-11	Sequence 11, Appl
c 14	347.5	6.6	712	4	US-09-270-767-1158	Sequence 1158, A
c 15	320	6.1	2119	4	US-09-620-312D-35	Sequence 35, Appl
c 16	291.5	5.5	7350	4	US-09-949-016-364	Sequence 364, Appl
c 17	284.5	5.4	3396	3	US-09-668-680-6	Sequence 6, Appl
c 18	280	5.3	3423	3	US-09-668-680-7	Sequence 7, Appl
c 19	279	5.3	213	4	US-09-513-999C-22310	Sequence 22310, A
c 20	268	5.1	6555	4	US-09-949-016-5295	Sequence 5295, Ap
c 21	261.5	5.0	2742	4	US-09-248-796A-1631	Sequence 1631, Ap
c 22	229.5	4.4	502	4	US-09-828-303-3	Sequence 3, Appl
c 23	229.5	4.4	2885	4	US-09-774-528-3	Sequence 3, Appl
c 24	223	4.2	1257	4	US-09-270-767-10482	Sequence 10482, A
c 25	221.5	4.2	5883	4	US-09-949-016-5001	Sequence 5001, Ap
c 26	221	4.2	4411529	3	US-09-103-840A-1	Sequence 1, Appl
c 27	219.5	4.2	4403765	3	US-09-103-840A-2	Sequence 2, Appl
c 28	217.5	4.1	1976	4	US-09-620-312D-515	Sequence 515, Appl
c 29	217	4.1	4378	2	US-09-080-897-3	Sequence 3, Appl
c 30	217	4.1	4378	3	US-09-323-735-3	Sequence 3, Appl
c 31	217	4.1	4399	3	US-08-899-595-2	Sequence 2, Appl
c 32	214.5	4.0	4403765	3	US-09-949-016-5411	Sequence 5411, Ap
c 33	210.5	4.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl
c 34	210.5	4.0	4411529	3	US-09-103-840A-2	Sequence 2, Appl
c 35	208.5	4.0	6202	4	US-09-774-528-120	Sequence 120, Appl
c 36	208	3.9	1698	4	US-09-599-287A-1	Sequence 1, Appl
c 37	206.5	3.9	39154	4	US-09-949-016-12384	Sequence 12384, A
c 38	206.5	3.9	39154	4	US-09-949-016-12801	Sequence 12801, A
c 39	206.5	3.9	39443	4	US-09-949-016-14326	Sequence 14326, A
c 40	206.5	3.9	39443	4	US-09-949-016-14327	Sequence 14327, A
c 41	206	3.9	1447	3	US-09-484-970B-121	Sequence 121, Appl
c 42	205.5	3.9	514	4	US-09-854-133-660	Sequence 660, Appl
c 43	205.5	3.9	229354	4	US-09-705-400-64	Sequence 64, Appl
c 44	204.5	3.9	8091	4	US-09-230-652-1	Sequence 1, Appl
c 45	204.5	3.9	8257	3	US-09-484-970B-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1

US-09-023-905A-8
; Sequence 8, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 2949
; TYPE: DNA
; ORGANISM: Danio rerio
; US-09-023-905A-8

Alignment Scores:
Pred. No.: 2.87e-249 Length: 2949
Score: 3874.00 Matches: 755
Percent Similarity: 83.12% Conservative: 92
Best Local Similarity: 74.09% Mismatches: 122
Query Match: 73.47% Indels: 50
DB: 4 Gaps: 13

US-09-914-042-1 (1-1006) x US-09-023-905A-8 (1-2949)

Qy 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 1 ATGGCTGACCAAGATAACAGTGGCGAGTTTGTACGGAGACAAATGAAGATTATAAATCG 60
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 61 CCCACCGCCTCAAACTTCACCAACAGAGTACTCACTGCAGGAAACACAGTATCCGCGCTG 120
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 121 GAGGAGGCCCTGGATGTGGACCGCAGTGTCTTTACAAGATCAAGATCAGTTAAGGCT 180
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 181 ATTTACGCCCTCGGGTCTCGCTCATGTGGAGAAATGAGGACGATCACTCAAGCTCTGGAG 240
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 241 AAGTTCCGAGAGAACTGTGTGTACAGAGATGACCCGGACCTGGGATCAGCCTTCTCGAAG 300
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 301 TTCTCCGCTCTTCCACCAAGAGCTCACGCGACTCTTCAAGAAACCTGTTTCAGAAACATGAAT 360
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 361 AATATATTACTCTCCCATTTGACAGTCTGCTGAAGGGAGATCTGAAAGGGGTAAAGGG 420
Qy 141 AspLeuLysLysPropheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle 160
Db 421 GATCTCAAGAAAGCCCTTCGATAAAGCTTGGAAAGACTACGAGACTAAAGTCTCTAAATA 480
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 481 GAGAGGAGAGAAAAAGACACGCCCGCGAGCAGCGAATGATCCGAGCGAGATCAGCGGA 540
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 541 GCAGAGATAGCAGAAAGATGGAANAAGCGCGGTTCCTCCAGCTTCAGATGTGTGAG 600
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 601 TACCTCTCAAAAGTCAATGAATCAAGATCAAAAAGGTGTGCGACTCTCCAGAAATCTC 660
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240
Db 661 ATCAAAATCTTCCACGCAAGTCAACTTCTTTACAGGATGGTCTCAAAAGCGGTGACAAAC 720
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 721 CTCAACCTCAATAGAAAAAATGCGCCACAGACTTGCACTCGATCAAAACAGGTACAGGAT 780
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 781 GAAGAACCACAGACAGCTAACCCAGTTACGGGATGTGCTAAAAACTGCTCTCAAGTGGAG 840
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 841 CAGAAGGAGGAGCTCTCAGGTAGACAGAGCGCCACCTACAGTCTGCACACCGCGCAGGCG 900
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 901 AACAAAGAGCATGGGACTGAGCGCAGCGCAACCTTTTACAAGAAAGTGAAGCGGCTCGCG 960
Qy 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 961 AAAGTGTGCAGAGAGAGAAAGTGCACAGTAAAGATGATATTTGACCATCTCCATCGG 1020
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1021 ACGGCAAAACAGACCTCCCGCCAAACTCAATCTTCTCACCTGTGAGGTAAGCAACCCA 1080
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380

Db 1081 GAGGAGAGAAAAAGTTTTTGACCTCATCTCATGACAGAACATATCATTTCCAGGCAGAA 1140
Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1141 GATGAGCAGAGTGTCAAAATATGATCTCAGTGTCTGCAGAACAGTAAAGAGAGCGCGTC 1200
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnThrGlyGluAsnIleValGlnGluLeuThr 420
Db 1201 AACAAACCCCTTCAAGGCGCACAGCATGTTGGTGAATAATAACATTTGTCAGGAGCTCACC 1260
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1261 AAGCCCATCTCGGAGAGGTGAACGGATGGCGGGGAACGATGTCTGCTGCGACTCGCGT 1320
Qy 441 AlaProaspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1321 GCTCCGCGCCACATGGCTCTCCCAACCTGGGCATCTCTGACCTGCATCGAGTGTTCG 1380
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1381 GGGATCCACAGAGAGCTGGGCGTCCATTACTCCGAATCCAGTCCCTCACACTCGACGTC 1440
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1441 CTCAGCACTCCGAGCTCTTGTGCGCCAGAACCTGGGGAATGTCTGCTTCAATGAGATC 1500
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1501 ATGGAGGCTGTCTGACGGCAGAAAGATGTGATCAACCGAATCCAGCCAGTGACATGCAG 1560
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaAaLysLysHis 540
Db 1561 GCGAGGAGGAGCTTTATCATGTGCGCAATACACAGAAACCGCTTCGCTGTAAGAAGTGT 1620
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1621 CCAGACGCATGTGCGAAGCTGCACCGCTCTGTGATGCTGTGAAGGCCCGGACATTTTC 1680
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1681 TCTCTCATCCAGGCTATGTCTGAAGGAGTGGATCTGATGGAGCCCATTTCTCTGGCTAAT 1740
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 1741 GGACATGAACAAGGTGAGACGGCTTTCATCTGCCCGTGAGACTGGTGGACAGAACTTCC 1800
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 1801 CTACACATCATCGACTTCTCCACCAAAACAGTTTAAACCTGGATAGCAAAACGGCTAAA 1860
Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 1861 GGAAGCAGCTCTGTCATTTACTGTGCTGACGAGCAACACAGCAGTGTCTCAAACTGTCTG 1920
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 1921 CTCAGAGAAAGCCTCCATAGATATCGTAAATGAAGCTGGAGAGACCCCTTGGACATC 1980
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 1981 GCCAGGCGACTCAAAACATCTGCAGTGTGAGAACTGCTGGAACACAGGCTCTTTCAGGGAAG 2040
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2041 TTCAATGTCTATGTGATGTGAGTATGAGTGGAGACTTCAGCATGAAGACTTGGACGAG 2100
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2101 AGTGAAGATCTGGATGGAGAGTCAAGTCTCTTCAAGGCGGAGTGAAGCGGCC 2154
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740

Db 2155 ATCAGCTCTACACACCGGCGAGTAACCTCCCTTCAGCTGAGTCCAGCAGCAGCTGAGCGGA 2214
Qy 741 AspalalaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAen 760
Db 2215 GACGCTGAGAGCTGGTTAAAGACAAGCAACGC---TTTGTGCCAAACCTGGTCAACAAT 2271
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAaProSer 780
Db 2272 GAAACCTACGGGACCATCATTAACACACAGCTCAACC-----GTGAGCGCTG 2316
Qy 781 ThrThrSerAlaProProLeuProProArgAenValGlyLysValGlnThrAlaSerSer 800
Db 2317 TCCTCTCTCTCCACTCTACCCACCCCGAACAATA-----GTGACGCTCTGCTCTT 2370
Qy 801 AlaaenThrLeu-----TrpLysThrAenSerValSerValAspGlyGly 815
Db 2371 GCAGGACTGACTCAAGGATCTCCCGGCTGGAAGCTGGCTCTCTGGATCTGAGCGGC--- 2427
Qy 816 SerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLeuArg 835
Db 2428 ---AGACAGAGATCTCTCTGACCTCCCAACATGATCTCTCTGCGCTCTCTTACGG 2484
Qy 836 ValThrSerThrAenProLeuThrProThr---ProProProValAlaLysThrPro 854
Db 2485 GTCACTTCCACTCTCTTAATGCCAGCGGTGCTGCTCTCTCTGCTGCTAAAGCTACT 2544
Qy 855 SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProProGlyIleSer 872
Db 2545 GGTATGAGGAGACCATGAATATGCACCCAAACCCGACAGGGGCTCTCTGGACAGAAC 2604
Qy 873 GlnIleArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysPro 892
Db 2605 ATCAACCGG----- 2613
Qy 893 AlaProGlyThrAsp-----LysSerThrProLeuThrAenLysGlyGlnPro 908
Db 2614 GCTCAAGTGGCGGCAAAAACCTCAGCAAAAGCACACTGATGCGCTCC----- 2661
Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927
Db 2662 ---GGATCCATCGAGACAGCAGCTAAAGATCCAGGAGGCGCCCAAAAACACCACTGGT 2718
Qy 928 ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys 947
Db 2719 CAAACTCTGCTGCGACCCACATGCCAGGAAA-----ACGTATTTGAAGCCGAAG 2769
Qy 948 ArgValLysAlaLeuTyrAsnCysValAlaAspAenProAspGluLeuThrPheSerGlu 967
Db 2770 CGTGTGAAGGCGCATGTATTAACGTGTGCGCGGATATATCCAGACGAGCTGACCTTCTGTAG 2829
Qy 968 GlyAspValIleIleValAspGlyGluGluAspGlnGluTrpTptIleGlyHisIleAsp 987
Db 2830 GGAGAGCTTATCGTGTGATGGAGAGAGAGACAGAGAGTGTGGTGGGCGACATTGAG 2889
Qy 988 GlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 2890 GGAGAGCAATAGAGAGAGGCGGTTCTCTGTCAGCTTTGTACAGTTTCATTATGAC 2946

RESULT 2

US-09-023-905A-6
; Sequence 6, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A

; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 5954
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (433) ... (3378)
US-09-023-905A-6

Alignment Scores:

Pred. No.: 7.5e-249 Length: 5954
Score: 3874.00 Matches: 755
Percent Similarity: 83.12% Conservative: 92
Best Local Similarity: 74.09% Mismatch: 122
Query Match: 73.47% Indels: 50
DB: 4 Gaps: 13

US-09-914-042-1 (1-1006) x US-09-023-905A-6 (1-5954)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleAla 20
Db 433 ATGCCTCACCAGATAACAGTGGCGAGCTTTGTCCAGGAGACAAATGAAGATTATAAATCG 492
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAenThrValAlaAlaIle 40
Db 493 CCCACCCTCCAAACTTCCACCACCAAGATGACTCCTGACGAAACACAGTATCCGACTG 552
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 553 GAGGAGGCGCTGGATGGACCGCAGTCTCTTTACAAGATGAAGAGTCAAGTAAAGSCT 612
Qy 61 IleAenSerSerGlyLeuAlaHisValGluAenGluGlnTyrThrGlnAlaLeuGlu 80
Db 613 ATTTACCGCTCGGCTCGGCTCATGTGAGAAATGAGGAGCAGTACACTCAAGCTCTGGAG 672
Qy 81 LysPheGlyGlyAenCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
Db 673 AAGTTCGAGAGAACTGTGTGTACAGAGATGACCCGACCTGGGATCAGCTTCTCTGAAG 732
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAenLeuIleGlnAenMetAsn 120
Db 733 TTCTCCGCTCTCACCAGGAGCTCAGCGCACTCTTCAAGAACCTGTTTCAGAACATGAAT 792
Qy 121 AsnIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 793 AATATCATTAACCTTCCCATTTGGACAGTCTGCTGAAGGAGAGATCTGAAAGGGGTTAAAGGG 852
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 853 GATCTCAAGAGCGCTTCGATAAAGCCTTGAAGAGCTACGAGACTAAAGTCTCTAAATA 912
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 913 GAGAGAGGAGAAAAGAGACGACCCCGCAGCAGCGAATGATCCGAGCGGAGATCAGCGGA 972
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 973 GCAGAGATAGCAGAAAGAGATGGAAGAGCGGGCTTTCTCCAGCTTCAGATGTGTGAG 1032
Qy 201 TyrLeuLysValAenGluIleLysIleLysLysGlyValAspLeuLeuGlnAenLeu 220
Db 1033 TACCTCTCAAGTCAATGAATCAAGATCAAAAGAGGTGTGACCTGCTCCAGATCTC 1092
Qy 221 IleLysTyrPheHisAlaGlnCysAenPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1093 ATCAATACTTCCACGACAGTGCACACTTCTTTCAGGATGTGTCTCAAGCGGTGACACAC 1152
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260


```
Db 1201 TGATGACGACATGATGAGAAATTCGACCCAGTCCCAACCGCGGGAGACCGGCCCAT 1260
Qy 721 eSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArgAs 741
Db 1261 CAGCTTCTACGAGCTGGGCTCCAAACGAGCTTCAGCTTAACGCTGTATCTTTGGCCAGAGA 1320
Qy 741 pAlaAlaSerLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsnG1 761
Db 1321 TGCTGCAAAACCTTGCCAAGGACAACGAGGAGCTTCATGCCCCAGCATCTTGCAGAAATGA 1380
Qy 761 uThrTyrGlyAlaLeuLeuSerGlySerProProProAlaGlnProAlaAlaProSerTh 781
Db 1381 GACTTACGAGCCCTCTCAGTGGGAGCCACCTCCCGCCAGCTCGAGCCCCAGCAC 1440
Qy 781 rThSerAla-ProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSerA 801
Db 1441 CACCAGCGCCCGCCCGCTTCCTCCACGGAATGTTGGCAAAAGTTCAGACAGCCTCCTCTG 1500
Qy 801 laAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSerS 821
Db 1501 CTAACACCCCTGTGGAAGACAAACTCTGTAAGTGTGAGCGGTGGAAGCCGCGAGCATCTT 1560
Qy 821 erSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsnP 841
Db 1561 CGTCAGATCCCGCAGCTGTCCATCCACCGCTGCCCTCTTCGCGTGACATCTACCAATC 1620
Qy 841 roLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeuS 861
Db 1621 CCCTGACCCCAACGCGCGCCCGCCACCGTTCGACAGACGCCAGCGTAATGGAAGCCTTGA 1680
Qy 861 erGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuProP 881
Db 1681 GCCAGCGGAGCAAGCTGCGCCGCTGGGATCTCACAGATCAGGCGCCCACTCTGCCCC 1740
Qy 881 roGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrP 901
Db 1741 CACAGCCCGCCCGCCCTCCCGCAGAGAGAGCTTCGCGCAGGGGCTGCAAGTCCACCC 1800
Qy 901 roLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuG 921
Db 1801 CACTGACCAACAAGGCCAACCGAGAGACCTGTGTGATCTCTCTGCAACGGAAGCTCTGG 1860
Qy 921 lyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGlnA 941
Db 1861 GTCCCTCTGTCCAATGCTATGGTCTTCGACGCCCCCTGCACCCCAATGCTAGGAAGTCGAGG 1920
Qy 941 laThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnProA 961
Db 1921 CAACCAAGTTGAGGCTTAAGCGGTGAAAGCGCTCTATAACTGTGTGCTGACAAACCCCG 1980
Qy 961 spGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGluGluAspGlnGluT 981
Db 1981 ATGAGCTCACCTTCTCCGAGGGGATGTGATCATCGTGACGGGAGGAGGACAGGAGT 2040
Qy 981 roTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPheV 1001
Db 2041 GGTGGATTGGCCACATGATGAGATCTCTGGTTCGCAAAAGGCGCATTCGCCGGTGTCAATTG 2100
Qy 1001 alHisPheIleAlaAsp 1006
Db 2101 TGCACCTTATCGCTGAC 2117
```

RESULT 4

US-08-322-742-12

; Sequence 12, Application US/08322742

; Patent No. 5688641

; GENERAL INFORMATION:

; APPLICANT: Sager, Ruth

; TITLE OF INVENTION: CANCER DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEES: Fish & Richardson

```
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322.742
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/938,823
; FILING DATE: September 1, 1992
; APPLICATION NUMBER: 07/844,296
; FILING DATE: February 28, 1992
; APPLICATION NUMBER: 07/552,216
; FILING DATE: February 28, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frazer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/048003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4328
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-322-742-12
```

Alignment Scores:
Pred. No.: 1,68-218 Length: 4328
Score: 3417.00 Matches: 648
Percent Similarity: 99.39% Conservative: 1
Best Local Similarity: 99.23% Mismatches: 3
Query Match: 64.80% Indels: 1
DB: 1 Gaps: 0

US-09-914-042-1 (1-1006) x US-08-322-742-12 (1-4328)

```
Qy 355 GlnValLysThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr 374
Db 3 CAGGTGAAGACCAACCCCTGAGGAGAGAGTGTGCTTACCTTATTTACATGACAGAACT 62
Qy 375 TyrHisPheGlnAlaGluAspGluGlnLysGlnIleTrpMetSerValLeuGlnAsn 394
Db 63 TACCACCTTCAAGCTGAAGATGAACAGGAATGTCAAAATATGATGTCTGTCTGCAAAAT 122
Qy 395 SerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsn 414
Db 123 AGCAAGAGAGAGCTTTAAACATGCAATTAAGGGGATGACATCTACGAGAGAAATAC 182
Qy 415 IleValGlnGluLeuThrLysGluIleLeuSerGluValGlnArgMetThrGlyAsnAsp 434
Db 183 ATCGTCCAAGAACTGACAAAGGAGATCATCTCAGAAGTGCAGAGATGACGGGCAATGAC 242
Qy 435 ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyLeu 454
Db 243 GTCTGCTGTGACTGTGGGGGCCAGATCTCATATGGCTTTCCACCAACCTGGGCATCTCG 302
Qy 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474
Db 303 ACCTGCATCGAGTGTTCGGAAATCCACGAGAGCTGGGGGTTCATTATTCAGGATGAG 362
Qy 475 SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsn 494
Db 363 TCCCTGACCTTAGATGTACTGGGAACATCTCAGCTGTCTGCTCGCAAGATATTGGGAAT 422
```

QY 495 AlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsn 514
Db GCAGGCTTAATGAGATCATGTAATGTTGCTTACCAGCTGAGGACTCAGTCAAAACCCCAAC 482
QY 515 ProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 534
Db CAGGCGGAGCATGAATGCAAGAAGGACTATACATCAGCCCAAGTATCATCGAGAGGAGA 542
QY 535 TyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVal 554
Db TAGCGCAAGGAGAACACCGGATAAACGGCGAAGCTTCACAGTCTTTGGAGGCCGCTC 602
QY 555 LysThrArgAspIlePheGlyLeuLeuGlnIleTyrAlaAspGlyValAspLeuThrGlu 574
Db AAAACGAGAGATATTTTGGATGCTCAAGCTTATGCTGATGCTGCTGCTTACGGAA 662
QY 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594
Db AAAATCCCACTGGCCCAACGGCATGAGCGGATGAACGGCCCTCCACTTGCAGTCA 722
QY 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeu 614
Db TCCGTGGATCGAACCTCTCTTACATTTAGACTTTTATGTTTTCAGAAACAGTGGNACCTG 782
QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634
Db GATAAACAGACAGGGAAGGAGCAGACAGCCCTGCACTACTGCTGCTGACCGACAATGCC 842
QY 635 GluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654
Db GAGTGCCCTCAAGTTGCTCTCGGGGGGAGGCTCCATCGAGATAGCAAAATGAGTCAGGA 902
QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThr 674
Db GAGACTCCGCTGGACATTTGCCAAGGCGCTCAAGACAGAGCACTGTGAGGAGCTGCTGACC 962
QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluThrArgLeuLeu 694
Db CAAGCCTTATCTGGAAGATTTAAATCTCACGTTACGTTGAATGAATGATGCGACTACTC 1022
QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerLeuAsn 714
Db CACGAAGACCTGGATGAAAGTGTACACATGATGAGAAATGACACCCAGTCCCAAC 1082
QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734
Db CGGCGGGAAGACCGGCCCATCAGCTTCTACAGCTGGGCTCCAACCACTTCAGTCTAAC 1142
QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754
Db GCTGTATCTTTGGCCAGAGATGCTGCAAACTTGGCAAGGACACAGCAGAGGCTTTTCATG 1202
QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774
Db CCCAGCATCTTGCAGAAATGAGACTTACGGAGCCCTCTCTGAGTGGCAGCCACCTCCGCC 1262
QY 775 GlnProAlaAlaProSerThrThrSerAla-ProProLeuProProArgAsnValGlyLys 794
Db CAGCTTGAGCCGCCAGCACCCAGCCGCCCTCCCTCCCTCCAGGAATGTGGCA 1322
QY 794 sValGlnThrAlaSerSerAlaAsnThrLeuThrLysThrAsnSerValAspGly 814
Db AGTTTCAGACAGCCTCTCTGCTAAACACCTGTGGAAGACAAACTCTGTAAGTGTGACGG 1382
QY 814 yGlySerArgGlnArgSerSerAspProProAlaValHisProProLeuProProLe 834
Db TGGAGCCGGCAGCGCATCTTCTCAGATCCCGCAGCTGTCCATCCACCGCTGCCCTCT 1442
QY 834 uArgValThrSerThrAsnProLeuThrProThrProProProProValAlaLysThrPr 854
Db TCGCGTGATCTACCAATCCCTGACCCCGCCGCGCCCGCCCGCCCGCTGCGCCCTCT 1502

QY 854 oSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLysSerGlnIle 874
Db CAGCGTAAATGAAGCTTTGAGCCAGCCGAGCAAGCTGCCCGCTGGGATCTCACAGAT 1562
QY 874 eArgProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPr 894
Db CAGGCCCCCCTCTGCCCCCAGCCGCCCTCCCGAAGAGAGCTTGGCC 1622
QY 894 oGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLe 914
Db AGGGCTGACAACTCCACCTGACCAACAAAGGCCAACCGAGAGACCTGTGATCT 1682
QY 914 uSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPr 934
Db CTCTGCAACGGAAGCTCTGGGTCTCTGTCCTGTCCTGTCGAGCCCTGTCACC 1742
QY 934 oMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAs 954
Db CATGCTTAGGAAGTGGCAGGCAACCAAGCTTAGCGGTGAAGCGCTCTATAA 1802
QY 954 nCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAs 974
Db CTGTGTGCTGACAAACCCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGGA 1862
QY 974 pGlyGluGluAspGlnGluThrTrpIleGlyHisIleAspGlyAspProGlyArgLysG 994
Db CGGGGAGGAGACGAGGAGTGTGATTTGGCCACATTTGATGAGATCTGTGTCGCAAGG 1922
QY 994 yAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db CGCATTCGGGTGTCTTTGTGCACCTTTATCGCTGAC 1959

RESULT 5

US-09-023-905A-11
; Sequence 11, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 2712
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-11

Alignment Scores:
Pred. No.: 2,65e-201 Length: 2712
Score: 3155.00 Matches: 621
Percent Similarity: 75.60% Conservative: 98
Best Local Similarity: 65.30% Mismatches: 174
Query Match: 59.83% Indels: 58
DB: 4 Gaps: 10

US-09-914-042-1 (1-1006) x US-09-023-905A-11 (1-2712)

QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 1 ATGCTCATCAGCTGACAGTACTGAGTTGTGATATTACCATGAGGACTATTAAGCA 60
QY 21 ProThrAlaSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaIle 40

Db CCGCAACATCAGTGTCTGCGCGCATGCGTCACTCAGGAATACAGTCGCGCTCTG 120
Qy GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db GAAGAGCGCTGGATCTGCGCGCATGCTGCTGCAAAAATGAAGATCAGTCAAGGCC 180
Qy IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
Db ATAAACAGCTCTGGTCAGACTCATGTAGAACAGAGGAGCAGTATCCAGGCCATAGAG 240
Qy LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db AGGTTTACGGATAACACTGTGTCAAAAGATGACCTCGAGATGTCCAATTAATCTCTACA 300
Qy PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db TTCGTGGTGTTCACCAAGAGGCTTACTGCTCTCTTTCAAGAACTTGCTACAGAACTGAAT 360
Qy AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db AACATCATCACTTTTCCACTAGACAGTCTGCTAAAGGGAGAGCCTCAAGAGGATCAAGGG 420
Qy AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db GATTTGAAAAGCCATTTGATAAAGCATGGAAGGATTAAGAAACCAACTGAGCAAGATT 480
Qy GluLysGluLysLysGluHisAlaLysLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db GAGAAAGAAAAGCAGAGACATGCCAAAACAGCAGCGTCTGATCCGAACAGAGATCAGTGA 540
Qy AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db GGAGAGATCGCAGAGAGATGGAGAAAGAGAGACCGCTGTTTCAGCTTCAGATGTGTGAG 600
Qy TyrLeuLeuLysValAsnGluIleLysLysLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db TACCTATTAAAGTGAATGAATCAAGTCAAAAAGGGGGTTCGACCTGCTTCCACACCTC 660
Qy IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db ATCAAAATCTTTTCATGCCCGAGTGCATTTCTTTCAGGATGGGCTAAAGGTCTGTGACAA 720
Qy LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db CTGAACCTTTTCATGAAAGCTTGCCACAGACTTAACCGCGAAACAAACAGACTCAAGAT 780
Qy GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db CGCAAAAGGAACAGTGTGTCAGCTGAAAGAACTCTTAATCTGCTCTACAGTCTGAG 840
Qy GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db TGTAAAGGAGGATGCTCAGTCAACAGCAGACGCGAGCTACAGTCTTCACCGAGTTGCGGGC 900
Qy AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db AATAAGGTCACGGCAGCGAGCGCTCTGGGATGCTCTCTCAACCGCAGCGAGGAGCTGAGG 960
Qy LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db AAAGTTTGGCAGAAAAGGAAGTCTCTGTGAAAATATGGATTGTTGACTATTTCACATGGA 1020
Qy ThrAlaAsnArgProProAlaLysLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db ACGCCCAATGCACCGCAGCAACCTGAACCTCTTAACCTGCGCAAGTGAAGCGTAACCCA 1080
Qy GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db GATGAGAAAAATATGCTTTGATCTCATATCATATGACAGAACGTATCATCTTCCAGACTGAG 1140
Qy AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400

Db GATGAGCGAGAGTGTCTGAGTATGGGTTTCTGTTCTCCAGAACAGTAAAGAGAGCGCGCTG 1200
Qy AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db AACAAATGCTTTAAAGACGATCAGAATGAGGGAGAAAATAACATTTGTTCCGAGAGCTCACT 1260
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db AAGGCCATCTGGGGGAAGTGAAGAAATGAGCGCAATGACGTGTGCTGTGCTGTGGA 1320
Qy AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db GCTTCCAATCAACATGGCTCTCCACAAACCTGGGTGTGTGATTGTCATTGAATGCTCTCT 1380
Qy GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuLeuAspVal 480
Db GGGATCCATCGGAAATGGCGCTCCACTACTCCCGAATACAGTCTCTGACACTGGACCTC 1440
Qy LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db TTAGGCACATCTGAACATTATGCTTAACAGTGTGGGAATGACGACATTCAATGAAATC 1500
Qy MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db ATGGAAGCAAACTGTCTTCAGAG--ATCCAAAACCCCTACCTTCTAGTGACATGCAG 1557
Qy AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db GTACGAAAGAACTTCATCACAGCCAAATACACAGAGAGCGTTTCGGCTCAGAAAGATAT 1617
Qy AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db GCATATACGAGCTCGACTGCATGCATGCTGTGTGATGCAGTGAAGTCTCGGACATCTTC 1677
Qy GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db TCCCTGATCCAGGCTATGCTGAAGGACTGACCTGTAGAGAGCAATTAATCAGCCTAAC 1737
Qy GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db CAACATGAACCAGCGAGACATCACTACATCTTGGGTACGAATGGTGGACCGAACTCC 1797
Qy LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db CTCATATTGTGGACTTCTTGACAGAACAGTGGCAATTTAGACAACGACAGCAGCCAAA 1857
Qy GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db GGAAGCACAGCGCTACATTTATGCTGCTTGAATACAGTGAATGATGAAGCTGCTG 1917
Qy LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db CTGCGGGGAAGGATCTGTGAGCATTAATGATCTGGAGAGACTGCTCTGGATTG 1977
Qy AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db GCGCAGCTCTCAAACTCCAAATGCGAGGAGTGTGCTGACTCAGGCGCAGCAGCGGAAG 2037
Qy PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db TTCAATGTCCATGTGCATGTGGAATATGACTGGGCTCTGTCATAATGAGGATCTGGACGAG 2097
Qy SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db AGCAAGATGATGAGAGGACAAG-----CCCATTCCTCCATCAGCGGTGAGAGGCTCCA 2151
Qy IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db ATAGCTGATATAGTTCAGGAGTGGCCCCCATGATGCCCAACATCAGCGCTCTGGCTCGG 2211
Qy AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db GACGTGGCCAAATGTGGTCAATAAAGCAGAGGGGCTTTTATTTCCGAGCATGATGAAC 2271

Db 1260 AAAAGTTTGGCAGAAAAAGGAAGTCTCTCTGTAAGAAATGGATTGTTGACTATTTCACATGGA 1319
Qy ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1320 AGCCCAATGACCCCGACCAAACTGAACTCTTAACTCGCAAGTGAAGCGTAAACCCA 1379
Qy GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1380 GATGAGAAAAAATGCTTTGATCTCATATCATACATGACAGACGATATCACTCTCCAGACTGAG 1439
Qy AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1440 GATGAGGAGAGTGTGAGTGTGCTGCTCTCTCCAGAAACAGTAAAGAAAGAGCGCTG 1499
Qy AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1500 AACAAATGCTTTAAAGACGATCAGAAATGAGGAGAGAAATAACATTGTTTCGAGAGCTCACT 1559
Qy LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1560 AAGGCCATCGTGGGGAGTGAAGAAATGAGCGCAATGACGTGTGCTGTGACTGTGGA 1619
Qy AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1620 GCTTCCAAATCCAAATGCTCTCCAAACCTGGGTGTGTTGATTGCAATTGAATGCTCT 1679
Qy GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1680 GGGATCCATCGGGAATGGGCTCCACTACTCCCGAATACAGTCTCTGACACTGGACCTC 1739
Qy LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1740 TTAGGCACATCTGAATATGCTTCTTAACAGTGTGGGAATGACGATTCAATGAATC 1799
Qy MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1800 ATGGAAGCAAACTGCTCTTCAGAG---ATCCCAAAACCTACCTCTCTAGTGACATGCAG 1856
Qy AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysIleHis 540
Db 1857 GTACGAAAGACTTTCATCAGACCAATACACAGAGAGCGTTCGCTCAGAAAGATAT 1916
Qy AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1917 GCAGATACCGAGCTCGACTGCATGCTGTGTGATGAGTGAAGTCTCGGACATCTTC 1976
Qy GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1977 TCCCTGATCCAGGCTATGCTGAAGGACTGGACCTGATGGAGACCATTAATCAGCCTAAC 2036
Qy GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2037 CAACATGAACCGGCGAGACATCATCTATCTTTCGGGTACGAATGTTGGACCGCAACTCC 2096
Qy LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2097 CTCCATATTGGGACTTCTGTACAGAACAGTGGCAATTTAGACAGACACAGCCCAA 2156
Qy GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2157 GGNAGCAGCGCTACATTATTGCTGCTTGAATGATAACAGTGAATGATGAAGCTGCTG 2216
Qy LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2217 CTGGGGGGAAGCATCTGTGACGATTTACTAATGATGCTGGAGAGACTGCTCTGATTTG 2276
Qy AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2277 GCGGAGCTCTCAACACATCCAAATGCCAGGAGCTGCTGACTCAGGCCACAGCGGGAG 2336
Qy PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db 2337 TTCAATGCTCCATGTGCAATATGATGCGCTCTGCAATTAATGAGGATCTGGACGAG 2396

Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
Db 2397 AGCGAAGATGAGTGGAGACAAG-----CCCATTTCCCATCAGCGGTGAGGAGGTCCA 2450
Qy IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2451 ATAAGCTGTATAGTTTCAGGCGAGTGGCCCATGATGCCAACATCAGGCGCTCTGGCTCG 2510
Qy AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2511 GACCTGGCCAATGGGTCAATAAAGCAGAGGCGCTTTATTTCGAGCATGATGTAAC 2570
Qy GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2571 GAGACTTACGGCACCATGCTGCATCCCACTCTCCACACTG-----GGTTTACCAGGA 2624
Qy ThrThrSerAlaProLeuProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2625 GTACCTGCGATTCCTCTTTTACCCCTCGGCCCTTGGNAGGGGA----- 2669
Qy AlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
Db 2670 -----TGG-----AGTCCCAATGGAGAACATCGGTAGACAGAGGTCA 2708
Qy SerSerAspProProAlaValHisProProLeuProLeuArgValThrSerThrAsn 840
Db 2709 TGTTCAATCTGCA-----AACCTCAAACTCTGNAACAAATAACTCTGTGTAT 2759
Qy ProLeuThrProThrProProValAlaLysThrProSerValMetGluAlaLeu 860
Db 2760 GTTCTGCTCTGCTCTCTCCACCTCTCTGCA----- 2792
Qy SerGlnProSerLysProAlaProProGlyIleSerGlnIleArgProProProLeuPro 880
Db 2793 -----CCCAAGAGACCTCCACCTCCAGATCCAAAGGCCAGTCTTCTCTCCAGCAGCC 2846
Qy ProGlnProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
Db 2847 ACGGCTCT-----CTGACCA-----TCCGCA 2870
Qy ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
Db 2871 CCGCTCTTATTCCACCTGCTCTCTCTCAGGCCAGCGCT----- 2909
Qy GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 2910 -----GTAGTGGCCCTCGCACCAAGTTATGCCCACTTCGTCA 2945
Qy AlaThrLysLeuLysProLysArgValLysAla 951
Db 2946 CTGACTGATGTCAAAAGTCTGCTGCTTAAGCC 2978

RESULT 7

US-09-023-905A-1

; Sequence 1, Application US/09023905A

; Patent No. 6475778

; GENERAL INFORMATION:

; APPLICANT: Roberts, Thomas M.

; APPLICANT: King, Frederick J.

; APPLICANT: Harris, David F.

; APPLICANT: Hu, Erding

; APPLICANT: Spiegelman, Bruce

; APPLICANT: Chan, Joanne

; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses

; TITLE OF INVENTION: Therefor

; FILE REFERENCE: DFN-021

; CURRENT APPLICATION NUMBER: US/09/023,905A

; CURRENT FILING DATE: 1998-02-13

; PRIOR FILING DATE: 1997-02-14

; NUMBER OF SEQ ID NOS: 39

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1
 ; LENGTH: 5330
 ; TYPE: DNA
 ; ORGANISM: Bos sp.
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (209)...(3595)
 US-09-023-905A-1

Alignment Scores:

Pred. No.: 1,88e-191 Length: 5330
 Score: 2013.50 Matches: 527
 Percent Similarity: 66.67% Conservative: 129
 Best Local Similarity: 55.29% Mismatches: 225
 Query Match: 57.15% Indels: 153
 DB: 4 Gaps: 23

US-09-914-042-1 (1-1006) x US-09-023-905A-1 (1-5330)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20
 Db 269 ATCGCGGACCAAGATCTCCGTCCTCCGAGTTTCATCGCGGAGACACCGAGGACTTACAACTCG 328
 Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 Db 329 CCCACCACTCCAGCTTCACTACGCGGCTGCACAACTGCAGGAACACCGTCACTGCTGCTG 388
 Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60
 Db 389 GAGGAGGCTCTAGACCAAGATAGAACAGCTTACAGAAAGTTAAGAAAGTCTGTAAAGCA 448
 Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGlu 80
 Db 449 ATATACAATTCGGTCAAGACCATGTACAAATGAAGAAACTATGCGCAAGTTCTTGTAT 508
 Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
 Db 509 AAGTTGGGAGTAATTTTAAAGTCAGACAAACCCAGATCTTGGCACCCTTTTGTCAAG 568
 Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120
 Db 569 TTTTCTACGCTTACAAAGAACTGTCCACACTGTGTAATAATCTCTCCAGGCGCTGAGC 628
 Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGly 140
 Db 629 CACAATGTGATCTTCACTTGGATTCTTGTGAAGGAGACCTTGAAGGAGTCAAAAGC 688
 Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
 Db 689 GATCTCAAGAAACCAATTTGACAAAGCTTGAAGAGATTATGAGACCGAAAGTTTACCAAAAT 748
 Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
 Db 749 GAGAAAGGAGAGAGGAGGAGCAGCCCAAGCAGACGCGGATGATCCGACCGAGATCAACCGC 808
 Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
 Db 809 GCCGAGATCGCGGAGGAATGAAAGGAGGCGGCGCTTCCAGCTCCAGATGTGCGAG 868
 Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuGlnAsnLeu 220
 Db 869 TATCTCATTAAGTAAATGAATCAAGACCAAAAGGCTGTGGATCTGCTGCAGAACCTG 928
 Qy 221 IleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGluSer 240
 Db 929 ATAAAGTATATACGACAGTGAATTTCTTTCAGATGGTTTGAAGACAGCTGATAA 988
 Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
 Db 989 TTGAACAGTACATTGAAAGCTGCTGCTGATTTGTATATATATCAACAGACCCAGGAC 1048
 Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280

Db 1049 GAAGAAAGAAACAGCTGACCGCAGCTCCGAGACCTAATAAAGTCTCGCTCCAACTCGAT 1108
 Qy 281 GlnLysGlu-----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGln 297
 Db 1109 CAGAAGGAGTCTAGGAGAGATTCCAGAGCGGAGGGA---GGCTACAGCATGCACCAG 1165
 Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
 Db 1166 CTGCAAGGCAACAAGGAATACGCGCAGCAGGAAGGGCTACCTGCTGAAGAAGAGTGAC 1225
 Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
 Db 1226 GGGATCCGGAAGGTGTGCGAGAGGAGGTCTCCGTCAGAACGGGATCTCTGACCATC 1285
 Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
 Db 1286 TCCCAAGCCACGTCACACAGACAGCAGCAAGCTGNAACCTTCTCACTTGCCAGGTGAAG 1345
 Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
 Db 1346 CCGAATGCGGAGGACAGAGTCTTTTGACCTGATATACATACAGGACGTATCACTTT 1405
 Qy 378 GlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
 Db 1406 CAGCGCGAAGATGAGCAGGATTATGACGTGGGATCTCGGTGCTGACAAACAGCAAGAG 1465
 Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln 417
 Db 1466 GAGGCCCTCACCATGGCTTCCGGGGGAAACAGAGTGTGGGGAGCAGCCTG---GAG 1522
 Qy 418 GluLeuThrLysGluIleSerGluValGlnArgMetThrGlyAsnAspValCysCys 437
 Db 1523 GAGCTGACGAAGGCCATCATCGAGACGTGACGGGCTCCCGGGCAACGACGTCTGCTGC 1582
 Qy 438 AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle 457
 Db 1583 GACTGCGCTCGGAGAACCCACCTGCTGCTCCACCAACTTGGGCATCTTGACCTGTATA 1642
 Qy 458 GluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThr 477
 Db 1643 GAATGTTCCGCGATCCATAGAGAAATGGGGTTCATATTTCTCGCATCCAGTCTTTGGA 1702
 Qy 478 LeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe 497
 Db 1703 CTAGACAAATTAGGAATCTTGAATCTTGTGCGCAAGAAATGTAGGAACAATAAGTTT 1762
 Qy 498 AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer 517
 Db 1763 AATGATATTATGGAAGCAAAATTTTACCCAGT---CCCTCACCACCAACCCACCCCTTCAAGT 1819
 Qy 518 AspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArg 537
 Db 1820 GATATGCTGTACGAGGAGATATATCTACGCAAGTATGTAGATCATAGTTTCACGG 1879
 Qy 538 LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg 557
 Db 1880 AAGACCTTTCATCGTCATCAGTAAACTGAACGAATTTGCTTGAGGCCATCAAAATCCAGG 1939
 Qy 558 AspilePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIlePro 577
 Db 1940 GATTACTTGCATTAATCAAGTCTATGCAAGGGGGTGGAGCTAATGGAACCCGCTG--- 1996
 Qy 578 LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp 597
 Db 1997 CTGGNACCGGACAGGAGCTTGGGAGACAGCCCTTCATCTTGCATGTCGAAACCGCAGAC 2056
 Qy 598 ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln 617
 Db 2057 CAGACATCTCTCCATTTGGTGGATCTTCTGTACAAAACTGTGGGAACCTAGATAAGCAG 2116
 Qy 618 ThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeu 637
 Db 2117 ACGGCCCTGGGGAACACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2176

```
Qy 638 LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro 657
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2177 AAGTGTGCTCAGAGCAGCCCACTGTGACGTCGTTATCAGCTGGAGAGCCGCC 2236
Qy 658 LeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu 677
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2237 CTGGACATAGCAAGAGAGACTGAAGCCCACTCAGTGTGAAGACCTGCTTTCCAGCTAAA 2296
Qy 678 SerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgGluLeuHisGluAsp 697
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2297 TCTGGAAGTTCAATCCACACGTCACGTCGAATATGATGGAATCTTTCGACGAGGAG 2356
Qy 698 LeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGlu 717
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2357 ATGGATGAGCGGATGACGACTGTGATGACAAACCGACCCCATCAAGAGCGGCTCC 2416
Qy 718 AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSer 737
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2417 CCCGACGCGAGAGCTTCTGCCACTCTCCAGCATCTCCCCCGAGGAC-----AAG 2467
Qy 738 LeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIle 757
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2468 CTCTCACTGCCCGGCTTCAGCAGCGCAAGGACAAACGACACTC----- 2512
Qy 758 LeuGlnAsnGluThrTyrGlyAla-----LeuLeuSerGlySerPro 771
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2513 -----TCTACGGCGCCTTCACAAACGATCTCGTCTCCACAGACACAGAC 2560
Qy 772 ProProAlaGlnProAlaProSerThrThrSerAlaProProLeuProProArgAsn 791
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2561 TCACCCACGTCACCGATCGCA-----GAGGCGCCCGCTGCTGCCCTCCAGAAAC 2608
Qy 792 ValGlyLys-----ValGlnThrAlaSer 799
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2609 GCCAGAAAGGTCCACCTGGCCACCTTCAACACTCCCTCTAAGCACCCAGACCTCTAGT 2668
Qy 800 SerAlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArg 819
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2669 GGCAGCTCACCTCTCCAGNAGGGTCTCTCCCCACACCCCGGA---CACAAAGA 2725
Qy 820 SerSerSerAspProAlaValHisProProLeuProProLeuArg-----ValThr 837
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2726 ACCCTGTCTGACCTCCAGCCCACTACTCTACGGGCCCCCAACAAAGCGCAGTTCCT 2785
Qy 838 SerThrAsnProLeuThrProThrProProProValAlaLysThrProSerValMet 857
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2786 TGGGTAACGAGTGGTCCCTCATCGTCC-----AGTAAGACCAACGAACAAAGTTC 2836
Qy 858 GluAlaLeuSerGln----- 862
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2837 GAGGCGCTGTCCACGACTCGAGCACCAGTCTCTGCAAGACTGCACTTGTCCCAAGAGTT 2896
Qy 863 -----ProSerLysProAla----- 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2897 CTTCTCTAACTACCTCAGAAAGTGGCACTAAGGAAACACAGACCAAGCATCTCTCTCC 2956
Qy 868 -----ProGlyLys-----SerGlnIle----- 874
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2957 CTCGACAAAGCAAGCTCCCACTGAGATCTTCCAGAAAGTCTGCCAGTTGACAGAGTTA 3016
Qy 875 -----ArgProProPro-----LeuProProGlnPro----- 883
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3017 CGCGAAGACCCCAACCGGGGACTGCCCCCGAAGCCCAACCGAACTGGCTCCCAACCC 3076
Qy 884 -----ProSerArgLeuProGlnLys----- 890
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3077 CCCATTGGAGACTTACCACCTAAGCCAGCGAGCTGCCCCCGAAGCCACAGCTGGCGCAG 3136
Qy 890 ----- 890
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3137 CTGCCCCCAAGCCCGAGCTCGCAGACTTGTCCCCCAAGCCCGCAGGTGAAAGACCTGCT 3196
```

```
Qy 891 ---LysProAlaProGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArg 909
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3197 CCCAAGCCCAACTGGGGAGCTGCTGCAGAAAAACCCAGAGCGGAGCGCTCCGCCAAG 3256
Qy 910 Gly-----ProValAspLeuSer----- 915
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3257 GCCCAGCCACCCCTGGAGCTCACCCCAAGTCAACCCGGCGGACCTGTCTCCCGAACGTC 3316
Qy 916 -----AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProPro 932
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3317 CCCAAGCAGGCGTCTGAGGACCAACAGACTCTCAGCCACCCTGCCA---GAGACACCC 3373
Qy 933 AlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeu 952
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3374 GTGCTCTGCCCGAAGATCAACACGGGAAGAGCAAGGTGAGCGGAGTGAAGACCATC 3433
Qy 953 TyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIle 972
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3434 TACGACTGCCAGCGGCAACAGATGACGAGCTGCTTTCATGGAGGGCGAGGTGATCGTG 3493
Qy 973 ValAspGlyGluGluAspGlnGluTyrTrpIleGlyHisIleAspGlyAspProGlyArg 992
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3494 GTCAACGGGAGGAGGACCAAGAGTGTGGATTGGGCATCGAGGGGACCCCGAGAGG 3553
Qy 993 LysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3554 AAGGCGCTCTCCAGTGTCTTTGTCCACATCCTGTCCGAC 3595

RESULT 8
US-09-023-905A-5
; Sequence 5, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; CURRENT FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3456
; TYPE: DNA
; ORGANISM: Danio rerio
US-09-023-905A-5

Alignment Scores:
Pred. No.: 9,06e-188 Length: 3456
Score: 2954.50 Matches: 622
Percent Similarity: 55.04% Conservative: 126
Best Local Similarity: 54.09% Mismatches: 239
Query Match: 56.03% Indels: 163
DB: 4 Gaps: 24

US-09-914-042-1 (1-1006) x US-09-023-905A-5 (1-3456)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 ATGCCGATCATGATCTCCGTGTCGAGTTTCTCTCGGAGACGACGAGGATTACAAATTC 120
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaIle 40
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 CCCAGACCTCGAGCTTCACCCCGCTCGAGAGCTGCCCGACACACCGTCAATGTTCTG 180
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysValLysAla 60
```

Db 181 GAAAGAGCTTTGGATCAGACCGCAACTCTTTACAGAAGGTCAGAAATCTGTCAAAGCA 240
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 241 ATCTACAACTCGGGTCAAGAACATGTGCAGAAATGAAGAGATTTATGGACAGGCACTGGAC 300
Qy 81 LysPheGlyGlyAsnCyValCyArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 301 AAGTTTGGCAGCAACTTCACTGAGCGAGATAACTCTGATCTGGGAACAGCCTTTCATCAAG 360
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLeuGlnAsnMetAsn 120
Db 361 TTTTCTGGACCTTATCAAAGAGCTGCTCTCTCTCAAGAAGCCTGCTCCAGAGCCTCAGC 420
Qy 121 AsnIleLeSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 421 CACAACGTCATCTTCAACCTGAGCTCTCTCTCAAGAGGAGATCTAAAGGGAGTGAAGGGG 480
Qy 141 AspLeuLysLysProPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLysIle 160
Db 481 GACCTTAAAGAGCCTTTCGCAAGGCTTGAAGAGACTATGAACCAAGTTCAAAAAGATC 540
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 541 GAGAAGGAGAGAGAGAACATGCCAAGCAGCACGGCATGATCCGCACAGAAATCACCGGC 600
Qy 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
Db 601 GCAGAGATTGCAGAGAGATGGAGNAGGAGCGGAGGATCTTTCAGCTGCAGATGTGTGAG 660
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 661 TACCTGATCAAGTCAATGAGATTAAAGACCAAGAGGAGTGGATCTCTCCAGAAATCTC 720
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 721 ATCAAGTATTATCATGCAGTGCATTTCTCCAGGATGGCTTGAAACCTGCTGACAAAG 780
Qy 241 LeuLysProSerIleGluThrSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 781 TTGAAGCAGTATATTGAATAATGACGCTGATCTTTATATATATATAAACAAGACTCAGAT 840
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 841 GAGGAGAAAAACAGCTCACAGCTCTCAGAGACCTCATCAATCTTCTTACAGCTGGAC 900
Qy 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 901 CAGAAGGAGGATTCTCAGAGTAAGCAGAGC---GGGTACAGCATGCACCAGCTGCAGGCG 957
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 958 AATAAGGAGTTTGGCAGTGAGAAGAGGCTATCTCTTCAAGAAGATGATGGATCCGT 1017
Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1018 AAGGTGTGCAGAGGAGGAGTGTCTCAGTGAATAATGTCATCTCTCAACCATCTCTCATGCC 1077
Qy 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
Db 1078 ACATCCAACAGCAGCGGTGAGACTGAATCTGCTGACCTGCAGCTGTCAGGTTAAACCCAGTGG 1137
Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1138 GAGGATAAGAAGTGTCTTGACCTCATCTCTCATATCAAGCAATATCATTTTCCAGGACAG 1197
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1198 GACGAAACAGGAGTTTGTGATGATCTCGGTGCTGACTAATAGTAAGGAGGAGGCTCTG 1257
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420

Db 1258 AACATGGCACTTTCGTGGGAGCAGAGTCTGGAGATGACAGCTTTG---GAGACTTGACC 1314
Qy 421 LysGluIleLeSerGluValGlnArgMetThrGlyAsnAspValCyCyAspCysGly 440
Db 1315 AAAGCCATCATCGAGGAGCTGCTGGCATCTCTGGAAACGAAGTCTGCTGTGACTGTGGG 1374
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1375 GTTCAGAGGCCCAATAGTTATCCATCAACCTCGGCATCTCTGACGTGATCGAGTGTCA 1434
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1435 GGATTCACAGGGAATGGAGTGCATATTTTCGGCATCCCAATCCATCGAGCTTGACAAA 1494
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1495 CTTGGAACCTCTGAACCTCTGCTGGCTTAAGAACGTGGGCAACAGTAGTTTCAACGAAATA 1554
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1555 TTAGAAGGGAATCTGCCGAGT---CCTTCAACAAGCCAGCGCCATCAAGTGACATGACC 1611
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrArgTyrAlaArgLysHis 540
Db 1612 GAGAGGAGGAGTACATCAATCGAAGTACGTGAGCACAGTTCGCTCGGCAACGCGC 1671
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1672 ACTACAGCCACAGCCAGACAGCGGACCTTGTACAGGCGGTGAGAACCGAGACTTGATG 1731
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1732 GCTCTCATTCAGCTCTATGCAGATGGATGGAGCTAATGGATCTCTTCCCAAGACA--- 1788
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 1789 GGACAGGACCGGAGAGACAGCTCTGCACCTTGTCTGGACATCATCAGACAGACTTCC 1848
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 1849 CTGCACCTGTGGACTTCTTGTCCAAAACAGTGGGACTCTTAGACAGACAGCGAGAGT 1908
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 1909 GGAACCGTGTCTCTCATCTTCTGTCACATATGAGAACCCAGAGTGTCTCAAACTGTGTG 1968
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 1969 CTCAGGGGAAAACCGTCTATTGACCTGTTAATCAAAACGGGGAGACAGCATTTGATATC 2028
Qy 661 AlalysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2029 GCCAGACGCTGAGNAATGTACAGTGTGAAGAGCTACTTGGTGGAGGAGCAGCGGAGG 2088
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db 2089 TTTAATCTCATGTGCATGTGGATGATGATGGAATCTCGCGCTGGAGAGATTGATGAG 2148
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2149 AGTGACGATACCTGGATGACAAGCTTAGTCCAGTGAAGAAGGAGCGTCTCTCGTCTCT 2208
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2209 CAGAGCTTC-----TGTCATTCTGTCAGCGTGTCT---CCTCAG 2244
Qy 741 AspAlaAlaAsnLeuAla-----LysGluLysGlnArg----- 751
Db 2245 GAGAAGTTAAACCTGCCGGGTATCTAGGACACAGGGAACAAGCAGAGACTGTCTCTATGGA 2304
Qy 752 AlaPheMet----- 754
Db 2305 GCCTTTGCCAACCCCGTCTACAGCACCTCCACCGAACCCTCGCATCTCCAGTGTGACAG 2364

```
Qy 755 ---ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPro----- 771
Db 2365 GGACCCACCATAGCCAGCAGACCCCTGCAAAAGCTCCGTCCTGTGGCGCCGCCACCTCT 2424
Qy 771 ----- 771
Db 2425 CTGCGCGCTGGGATCTCAATCGAGTCAGGAGGCGAGCTCCACTTCTGTCTAAGAGAGAGCT 2484
Qy 772 ProProAlaGlnProAlaProSerThrThrSerAlaProProLeuPro----- 788
Db 2485 CTTCTCCACCTCCCGGACACAAAGGCGACCCACTCAGATCCGCCCGCAGTCCGCTACTGCGAG 2544
Qy 789 ---ProArgAsnValGlyValGlnThrAlaSerSerAlaAsnThrLeuTriPlysThr 807
Db 2545 GGTCCCGAGAGAAAGAGTGTAGTCCACCTCTCTGCAATCGGACATCCCCCGGCC 2604
Qy 808 AsnSerValSerValAspGlyGlySerArgGlnArgSerSerSerAsp----- 823
Db 2605 AAC-----AAGTTTGAGGGAATCCAGCAGCAGCAAGCACTACGTCTATGNACACAAAA 2658
Qy 824 -----ProProAlaValHisProProLeu----- 831
Db 2659 GCAACATTTGGCCCGCCAGGATTTCTCCCAAACTACCTCAAAAGTGGCACTACGAAAGATT 2718
Qy 831 ----- 831
Db 2719 GACAAATCCACCTCCCATCATAGTGACAAAGTCTGGTCTGTGATGTCTTCAAGAACCCCA 2778
Qy 832 -----ProProLeuArgValThrSerThrAsnProLeuThr---ProThr 845
Db 2779 CAGGCCAGGATGCACCTCCACAGAGCCTCAGATACA-----ATAACCCAGACCCACT 2832
Qy 846 ProProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2833 GAACCTCCACCT-----AAAATTCACAGGTGCGAGAACGATCCAGCCTGTGTGATGTC 2886
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProPro-----LeuProPro 881
Db 2887 CCGCAGAACCGCACATCTCAGACCTTCTCCCAAAACCGCACTATCAGATCTTCCCGCCC 2946
Qy 882 GlnPro-----ProSerArgLeuProGlnLysLysProAlaProGlyThr 896
Db 2947 AAACCCCAATTCTCGGATTTACCACCAAAA---CTCAGCTTCTGTGACTGCCCGCGAG 3003
Qy 897 AspLysSerThrProLeuThrAsnLysGlyGln-----ProArgGlyProVal 912
Db 3004 CTTCAGCTTAAGGATCTTCCCGCTTAAGCGCGAGATCAGTGATCTGCCATCCAAACCGGCC 3063
Qy 913 AspLeuSerAlaThrGluAla----- 919
Db 3064 GTGTGTTCTGCGTCTGAGGCGCACACAGAGCGCATCAACGAGGAGGAACACAGTCCGAAG 3123
Qy 920 -----LeuGlyProLeu----- 923
Db 3124 CCCAGCTGACGAGACACAGTCATTCAGCCAGCAGGAGGAGCTCTCACCCCGCAGGCC 3183
Qy 924 -----SerAsnAlaMetValLeuGlnProProAlaProMetProArg 937
Db 3184 AGCAGGACACCAATGGAGCGCGCCGAGGAGCTTTGGAAATGCCAGTCCCAATGCCACGC 3243
Qy 938 Lys---SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysVal 956
Db 3244 AAAATTAAACAGATGACAAAGAACAAACGAGCGGTGTGAAMACCATCTATGATGGCCAG 3303
Qy 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGlu 976
Db 3304 CGAGACATGAGATGAGTGACTTTTGTGGGGCGGAGGTATTAATGTTCACAGAGAG 3363
Qy 977 GluAspGlnGluTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996
Db 3364 GAAGACCGAGGATGTGGTATCGGCACATAGAGGCTCAGCCTGAAGAAAGGGGTCTTTC 3423
```

```
Qy 997 ProValSerPheValHisPheIleAlaAsp 1006
Db 3424 CCATGTCTCTCGTGACATCTCTGTCTGAC 3453
RESULT 9
US-09-023-905A-3
; Sequence 3, Application US/09023905A
; Patent No. 6475778
; GENERAL INFORMATION:
; APPLICANT: Roberts, Thomas M.
; APPLICANT: King, Frederick J.
; APPLICANT: Harris, David F.
; APPLICANT: Hu, Erding
; APPLICANT: Spiegelman, Bruce
; APPLICANT: Chan, Joanne
; TITLE OF INVENTION: Differentiation Enhancing Factors and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: DFN-021
; CURRENT APPLICATION NUMBER: US/09/023,905A
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/038,191
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 4382
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (351)...(3803)
US-09-023-905A-3
Alignment Scores:
Pred. No.: 1,25e-187 Length: 4382
Score: 2954.50 Matches: 622
Percent Similarity: 65.04% Conservative: 126
Best Local Similarity: 54.09% Mismatches: 239
Query Match: 56.03% Indels: 163
DB: 4 Gaps: 24
US-09-914-042-1 (1-1006) x US-09-023-905A-3 (1-4382)
Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 411 ATGCCGATCAGATCTCGTGTCCGAGTTCTCTCGGAGACGCGGAGGATTACAATCC 470
Qy 21 ProThrAlaSerSerPheThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 471 CCCAGACCTCGAGCTTCACCACCCGCTGCGAGAGCTGCCGGAACACACGGTCAATGTTCTG 530
Qy 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLysAla 60
Db 531 GAAGAGGCTTTGGATCAGGACCGAATCTCTTACAGAGGTCAAGAAATCTGTCAAAGCA 590
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 591 ATCTACAACTCGGTCAAGAACATGTGAGATGAGAGAAATTATGGACGCGACTGGAC 650
Qy 81 LysPheGlyGlyAsnCysValCysArgAspAspProAspLeuGlySerAlaPheLeuLys 100
Db 651 AAGTTTGGCACTTCATCAGCCGAGATACTCTGATCTGGGAACAGCTTCATCAAG 710
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 711 TTTTCTGAGCTTATCAAGAGCTGGCTCTCTCTCAAGAACCTGCTCCAGAGCTCAGC 770
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
Db 771 CACAACTGCTCTTCAACCTGGACTCTCTGCTCAAGAGGAGATCTAAAGGGAGTGAAGGG 830
Qy 141 AspLeuLysLysProPheAspLysAlaTriPlysAspTyrGluThrLysIleThrLysIle 160
```

831	Db	GACCTTAAAGACCTTTTCGACACAGGCTTGGAAAGACTATGATAAAACCAAGTTCACAAAGATC	890
161	Qy	GLuLySGluLySLysGluHiSAlaLySLeuLeuHiSGlyMetIleArGThrGluIleSerGly	180
891	Db	GAGAAAGAGAGAGAGAGAACATGCGCAAGCAGCAGCGCATGATCCGACACAGAAATCACCGGC	950
181	Qy	AlaGluIleAlaGluGluMetGluLySGluArgPhePheGlnLeuGlnMetCysGlu	200
951	Db	GCAGAGATTGCAGAAAGAGATGGAGAGAGCGGAGGATCTTTTCAGCTGCAGATGTGTGAG	1010
201	Qy	TyrLeuLeuLysValAsnGluIleLysIleLysLySGlyValAspLeuLeuGlnAsnLeu	220
1011	Db	TACCTGATCAAAGTCAATAGATTAAGACCAAGAAAGGAGTGGATCTCTCCAGAAATCTC	1070
221	Qy	IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer	240
1071	Db	ATCAAGTATTATCATGTCACAGTGCAAATTTCTTCCAGATGGCTTGAAAACTGCTGACAAG	1130
241	Qy	LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp	260
1131	Db	TTGAAGCAGTATATTGAAAAATTAGCAGCTGATCTTTATAATAATAAAACAGACTCAGAT	1190
261	Qy	GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu	280
1191	Db	GAGGAGAAAAAACAGCTCACAGCTCTCAGAGACCTCATCAAAATCTCTTACAGCTGGAC	1250
281	Qy	GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly	300
1251	Db	CAGAAGGAGGNTTCTCAGATTAACACAGC---GGGTACAGCATGCACAGCTGCAGGCG	1307
301	Qy	AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg	320
1308	Db	AATAAGAGATTGGCAGTGAGAAAGAGGCTATCTCTTCAAGAAGAGTGATGGGATCCGT	1367
321	Qy	LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly	340
1368	Db	AAGGTGTGGCAGGAGGAAAGTGTCTAGTGAATAATGGCATCTCTCACCATCTCTCATGCC	1427
341	Qy	ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro	360
1428	Db	ACATCCAACAGGCAGCGGTGAGACTGAACTGCTGTGACCTGCCAGGTAAACCCAGTGGGA	1487
361	Qy	GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu	380
1488	Db	GAGATAAGAAGTCTTTTGACCTCATCTCTCATATAATCGAACATATCATTTCCAGGCAGAG	1547
381	Qy	AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu	400
1548	Db	GACAAACAGGAGTTTGTGATATGGATCTCGTGTCTGACTAATAGTAGAGGAGGCTCTG	1607
401	Qy	AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr	420
1608	Db	AACATGGCATTTCTGGGGACAGAGTGTGGAGATGACAGATTG---GAGGACTTGACC	1664
421	Qy	LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly	440
1665	Db	AAAGCCATCATCGAGGAGCTGTCGCGCATCTCTCGAAACGAAGTCTCTGTGACTGTGGG	1724
441	Qy	AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer	460
1725	Db	GTTCACAGGCCAAATGGTTATTCACATACTCGGCATCTCTGACGTGCATCGAGTGTTC	1784
461	Qy	GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal	480
1785	Db	GGAAATCCACAGGGAATGGGAGTCCATTTTCGGCATCCAAATCCATGGAGCTTGACAAA	1844
481	Qy	LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle	500
1845	Db	CTTGGAACTCTGAACCTTGTGCTTAAGAACCTGGGCAACAGTAGTCTTCAACGAATA	1904
501	Qy	MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn	520
1905	Db	TTAGAAGGGAATCTGGCGAGT---CCTTACCAAAGCAGCGCCATCAAGTACATCACC	1961

QY	521	AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis	540
DB	1962	GAGAGGAAGGAGTAGACATCAATGCGAAGTAGCTGGAGCACAGGTTGCGTCCGCGCAACGCC	2021
QY	541	AlaAspAenAlaLysLeuHisSerLeuCyAGluAlaValLysThrArgAspIlePhe	560
DB	2022	ACTACAGCCACGACGACGAGCGGCGACTTGTACAGGCGGTGAGAACGCCGAGACTTGTGATG	2081
QY	561	GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAen	580
DB	2082	GCTCTCATTACGCTCTATGCAGATGGAGTGGAGCTAATGGATCCTTTTCCGAGAACA	2138
QY	581	GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer	600
DB	2139	GGACAGACCCCGGAGAGACAGCTCTGCACTTTCTCGGACATCAGACACGAGACTTCC	2198
QY	601	LeuHisIleValAspPheLeuValGlnAenSerGlyAenLeuAspLysGlnThrGlyLys	620
DB	2199	CTGCACCTGGTGGACTTCTTGTCTCAAAACAGTGGGACTCTAGACACAGACGCGAGAGT	2258
QY	621	GlySerThrAlaLeuHisTyrCyAGLysLeuThrAspAenAlaGluCyLysLeuLysLeuLeu	640
DB	2259	GGAAACGCTGCTCTCCATTACTGCTGCATATGAGAGAGCCAGAGTGCTCTCAACTGCTG	2318
QY	641	LeuArgGlyLysAlaSerIleGluIleAlaAenGluSerGlyGluThrProLeuAspIle	660
DB	2319	CTCAGGGGAAACCGTCTATTGACCTGTTAATCAAAACGGGAGACAGCATTTGGATATC	2378
QY	661	AlaLysArgLeuLysHisGluHisCyAGluGluLeuLeuThrGlnAlaLeuSerGlyArg	680
DB	2379	GCCAGACACTGAGAAATGTACAGTGTGAAGAGCTACTGCTGGAGGACGAGCGCGGAGG	2438
QY	681	PheAenSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu	700
DB	2439	TTTTAATCTCATGTGCATGTGGAGTATGAGTGGAAATCTGGCGCTGGAGGAGATTGATGAG	2498
QY	701	SerAspAspMetAspGluLysLeuGlnProSerGluAenArgArgGluAspArgPro	720
DB	2499	AGTCAGCATGACCTGGATGACACGCTAGTCCAGTGAAGAAGGAGCGTCTCTCGTCCT	2558
QY	721	IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAenAlaValSerLeuAlaArg	740
DB	2559	CAGAGCTTTC-----TGTCATTCTGCTCCAGCGTGTCT---CCTCAG	2594
QY	741	AspAlaAlaAenLeuAla-----LysGluLysGlnArg-----	751
DB	2595	GAGAAGTTAACCTTCGCGGGGTATCTAGGACACAGGGACAGGACAGACTGTCTCATGGGA	2654
QY	752	AlaPheMet-----	754
DB	2655	GCTTTTGCCAACCCGCTCTACAGCACCTCCACGAAACCCCTGCATCTCCAGTGTCTCAGAG	2714
QY	755	---ProSerIleLeuGlnAenGluThrTyrGlyAlaLeuLeuSerGlySerPro-----	771
DB	2715	GGACCCACCATAGCAGCAAGACCCCTGCTCAAAAGCTCCGCTCTGTGGCGCGCCACCTCT	2774
QY	771	-----	771
DB	2775	CTGCGCGTGGGATCTCAATCGAGTGCAGGAGCGAGCTCCACTTTGTTCTAAGAAAGAGAGCT	2834
QY	772	ProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro-----	788
DB	2835	CCTCTCCACCTCCCGGACACACAGCGCACCCACTCAGATCCCCCCAGTCCCGTACTGCAG	2894
QY	789	---ProArgAenValGlyLysValGlnThrAlaSerSerAlaAenThrLeuThrPheThr	807
DB	2895	GGTCGCGAGACAAAGGAAGTAGTCCACACCTCTCTCTCAAAATCGGACATCCCCCGGCC	2954
QY	808	AenSerValSerValAspGlyGlySerArgGlnArgSerSerSerAep-----	823
DB	2955	AAC-----AAGTTTGAGGGAATCCAGCAGCAGCAAGACTACGTCTATGAACACAAAAA	3008

```
Qy 824 -----ProProAlaValHisProProLeu----- 831
Db 3009 GCAACATTTGGCCACGAGTTCTTCCAAACTACCTCAAAAGTGGCACTAGAAAGATT 3068
Qy 831 ----- 831
Db 3069 GACAAATCCACCTCCCATCAGTGACAAAGTCTGGTCTGTGATGTCTTCAGAAACCCCA 3128
Qy 832 -----ProProLeuArgValThrSerThrAsnProLeuThr---ProThr 845
Db 3129 CAGGCCAGGATGACACCTCCACCAAGCCCTCAGATACA-----ATAACAGACCCCACT 3182
Qy 846 ProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 3183 GAACCTCCACCT-----AAATTCACAGTTCGACAAACGATCCACGCTGTGTGATGTC 3236
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProPro-----LeuProPro 881
Db 3237 CCGCAGAAACCGCACATCTCAGACCTTCTCCAAACCGCAACTATCAGATCTTCCGCCCC 3296
Qy 882 GlnPro-----ProSerArgLeuProGlnLysLysProAlaProGlyThr 896
Db 3297 AAACCCCAATTTGTCGATTATACCAACCAAAA---CCTCAGCTTTCTGACCTGCCCCCGAAG 3353
Qy 897 AspLysSerThrProLeuThrAsnLysGlyGln-----ProArgGlyProVal 912
Db 3354 CCTCAGCTTAGGATCTTCCCTTAAGCGGAGATCAGTGATCTGCCATCCAAACCGGCC 3413
Qy 913 AspLeuSerAlaThrGluAla----- 919
Db 3414 GTGTGTTCTGCTGCTGAGGCCACACAGAGCGAGTCAACGACGAGGAAACCAAGTCCGAAG 3473
Qy 920 -----LeuGlyProLeu----- 923
Db 3474 CCCAGCTGACGGAGACACAGTCATTCAGCCAGCAGGAGGAGCTCTCACCCCGACAGGCC 3533
Qy 924 -----SerAsnAlaMetValLeuGlnProProAlaProMetProArg 937
Db 3534 AGCGAGGACCAATGGAGCGCCCGCAGGAGCTTGGAAATGCCAGTCCCAATGCCACGC 3593
Qy 938 Lys---SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyAsnCysVal 956
Db 3594 AAATAACACAGTAGCAAGAAAGAAACGAGCGTGTGAAACCATCTATGATGTCAG 3653
Qy 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGlu 976
Db 3654 GCAGACAAATGACGATGAGCTGACTTTTGTGGAGGCGGAGGTTATATATTTGTACAGGAG 3713
Qy 977 GluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996
Db 3714 GAAGACCAAGGAGTGGTGGATCGGCACATAGAGGTCAGCCTGAAAGAAAGGGGTCTTTC 3773
Qy 997 ProValSerPheValHisPheIleAlaAsp 1006
Db 3774 CCAATGCTCTTCGTGCACATTCGTGCAGAC 3803
```

RESULT 10

US-09-620-312D-155
; Sequence 155, Application US/09620312D
; Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John fillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes version 1.0
SEQ ID NO 155
LENGTH: 6990
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (61)..(2397)
US-09-620-312D-155

Alignment Scores:
Pred. No.: 3,28e-28 Length: 6990
Score: 571.50 Matches: 196
Percent Similarity: 41.68% Conservative: 137
Best Local Similarity: 24.53% Mismatches: 305
Query Match: 10.84% Indels: 161
DB: 4 Gaps: 25

US-09-914-042-1 (1-1006) x US-09-620-312D-155 (1-6990)

Qy 38 AlaAlaIleGluGluAlaLeu---AspValAspArgMetValLeuTyLysMetLysLys 56
Db 115 GCAGCTTTGGGAAGTAGAAGGTGATGGCAGCAATTTGGAACTA---AAACTTGATAAG 171
Qy 57 SerValLysAlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyThr 76
Db 172 CTTGTGAACCTT-----GGGATTGCAATGATTGATCTACTGGAAAGCCCTTTGT 219
Qy 77 GlnAlaLeuGluLysPheGlyAsnCysValCysArgAsp----- 90
Db 220 GTTCAATAAACAGTTCATCAATGGG-----ATTGAGACCTGGCCAGTATTTCTAGT 273
Qy 91 ---AspProAspLeuGlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThr 109
Db 274 AATGATGCTGCTGTGAGACAAAGTTTGACCAAGTTTCTGACAGTCTTCAAGAAATGATA 333
Qy 110 AlaLeuPheLysAsnLeuIleGlnAsnMetAsnIleIleSerPheProLeuAspSer 129
Db 334 AATTTTCACACAATCCTGTTGCCAAACTCAGAGATCAATTTTAAGGCACAGCTTCAGTAAC 393
Qy 130 LeuLeuLysGlyAspLeuLysGlyValLysGlyAspLeuLysLysProPheAspLysAla 149
Db 394 TTTGTTAAAGAAAGATCTTAGAAATTTCAA---GATGCCAAGAAAGCAATTTGAAAGATC 450
Qy 150 TrpLysAspTyThrGluThrLysIleThrLys-----IleGluLysGluLysGlu 166
Db 451 AGTGAGAAAGAAAGAAATCGCTTAGTAAAAAATGCCCAAGTACAAAGAAACAAACACAT 510
Qy 167 HisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIleAlaGlu 186
Db 511 GAAGTTGAA-----GAAGCCCAACCAAT 534
Qy 187 MetGluLysGluArgArgPhePheGlnLeuGlnMetCysGluTyThrLeuLysValAsn 206
Db 535 CTGACAGCAACAAAGAAATGTTTCCGACACATAGCCCTCGATTATGCTTCCAGATTAT 594
Qy 207 GluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeuIleLysTyPheHisAla 226

Db 595 GTTCTTCAATCAAAAGGAGATCAGAAATCTTAAATCAATGTTGTCTATTTATGTATGCC 654
 QY 227 GlnCysAsnPhelcAspGlyLeuLysAlaValGluSerLeuLysProSerIleGlu 246
 Db 655 CATTTGGCCCTTTTCATCAAGGATATGATCTGTTAGTGAATCTTGGACCTTACATGAAG 714
 QY 247 ThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluGluArgGlnLeu 266
 Db 715 GATCTTGTGGCAGTGTGATCGATCTGTTGGATGCGAGCAAGAGGAGAAAGAGAAATG 774
 QY 267 IleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGluGlnLysGluAspSerGln 286
 Db 775 GAGCAAAACATTC-----ACCATTCAACAAGGATTTCTCCAGT 816
 QY 287 IleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThr 306
 Db 817 GATGATTTCTAAGTTAGATATAACGTA-----GATGCTGCAAAATGGCATA 861
 QY 307 GluArgAsnGlySerLeuTyrLysSerAspGlyIleArgLysValTrpGlnLysArg 326
 Db 862 GTTATGGAAGGATATCTGTTCAACAGCGCCAGCAATGCCCTTCAAAACTTGGAAACAGCGCC 921
 QY 327 LysCysSerValLysAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProPro 346
 Db 922 TGGTTTTCAATACAGAAATCAATCAGTTGTTTACCAGAAAAAATTTAAGGATTAATCCGACT 981
 QY 347 AlaLysLeu---AsnLeuLeuThrCysGlnValLys-----ThrAsnProGluGluLys 363
 Db 982 GTGGTAGTTGAAGACCTCAGGCTTTGCACAGTGAACATCTTGAAGACATAGAGCGACA 1041
 QY 364 LysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGluAspGluGln 383
 Db 1042 TTCTGCTTTGAGTGCTCGCAACAAAGATTGCTCATGCTCCAGGCAGATTCGCAAAAG 1101
 QY 384 GluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeuAsnAla 403
 Db 1102 CTGCCAGGCGATGATTAAGCTGTTTCAGACCAGTATTTGCTACTGCTTATAGAGAG--- 1158
 QY 404 PheLysGlyAspAsp-----AsnThrGlyGlu--- 412
 Db 1159 ---AAGGTGATGATCAGAGAGCTGGATAAGAAATCATCTCCATCCACAGGAGCCTA 1215
 QY 413 -----AsnAlaValGlnGluLeuThrLys-----GluIleIleSerGluVal 427
 Db 1216 GATTCGGAATAGTCCAAAGAGAAATTAATTGAAGGAGAAAGTGCCTTCAGCGGCTC 1275
 QY 428 GlnArgMetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeu 447
 Db 1276 CAGTGATCTCCGCAATGCCAGCTGTTGTGACTGTGGCTGGCAGATCCAGGTGGGCC 1335
 QY 448 SerThrAsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGly 467
 Db 1336 AGCATCAACCTGGGCATCACCTTGTGTATCGAGTCTCCGGAATTCACGGAGCCTTGGG 1395
 QY 468 ValHisTyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeu 487
 Db 1396 GTTCATTTTCAAAAGTAGCATCTTTAACTTTAGACACTGGGAGCCAGAACTTTTAAAG 1455
 QY 488 LeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAla 507
 Db 1456 CTTATGTGTGAGTGGGAGATGATGTTATATAATCGAGTTTATGAAGCTTAATGTGGAATA 1515
 QY 508 GluAspSerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThr 527
 Db 1516 ATGGGAATAAGAAACCCCAACGAGA---CAAACACAGGAGGAGGAGGCATATATCAGA 1572
 QY 528 AlaLysTyrIleGluArgTyrAla----- 536
 Db 1573 GCAAAATATGTGAGAGGAAATTTGTGGATAAATATTCTATATCATTTATCACCTCCTCGAG 1632
 QY 537 ---ArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCys----- 551
 Db 1633 CAGCAAAAAAAGTTTGTCTCTTAAAGTTCTTGAAGAAAAAGAGGCTGAGCATTTCTAAATTT 1692

QY 552 -----GluAlaValLysThrArgAspIlePhe 560
 Db 1693 GGGCCAGGGGACCAAGTCAGATCTGCCCCAAGTTTCAGTCAGAAGTAATGAC---AGT 1749
 QY 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 1750 GGAATTTACAGACAGCTCTGATGATGAAGAGAATCTTTACCTCCACGGTGTACGCCAAT 1809
 QY 581 Gly---HisGluProAsp----- 585
 Db 1810 AGTTTATATGAGCTGAAGGAGAAAGGCAAGATTTCTTATGTTTCTTACTCGAAACAT 1869
 QY 585 ----- 585
 Db 1870 CTTAATCCAGACTTCAGCTTTATAGGGCGTCATATGAAAAAACCTTCTCTAAATGGCT 1929
 QY 586 GluThrAlaLeuHisLeuAla----- 592
 Db 1930 GAGGCTTTGGCTCATGCTGGTGACAGCTGAATCGGGCCAATTCAGAGGAAAAACAAGCGACA 1989
 QY 593 -----ValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsn 610
 Db 1990 CCACCTTATTCAGCTGTATTTAGGGGCTCTTTGGTGACGCTGTGAGTTCTCTCTACAGAT 2049
 QY 611 SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeu 630
 Db 2050 GGTGCTAATGTCAACCAAGAGATGTCCAAGGGCGGGACCATTCGCCATCCACCGCTC 2109
 QY 631 ThrAspAsnAlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluLeuAla 650
 Db 2110 TTAGGGCACACAGGCGAGTATGTTTATTCCTAAACGAGGTGCCAATCAACATGCCACT 2169
 QY 651 AsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGlu 670
 Db 2170 GATGAGAGGAGAAAGACCTTTTGACATAGCTGTGGAAGAGCGCAATTCATATATAGTC 2229
 QY 671 GluLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGlu 690
 Db 2230 ACCTTGTTCCTTAGCA----- 2247
 QY 691 TrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGln 710
 Db 2248 -----AGAATGAATGAAGAGATGCGGAATCAGAGGACTTTATGGA-----CAG 2292
 QY 711 ProSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
 Db 2293 CAGGTGATGAACCTTATCAGACATATTTCTGTGATTTTTCCTCAATGGCATCCTCAAT 2349

RESULT 11

US-09-270-767-1629/c
 ; Sequence 1629, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1629
 ; LENGTH: 1090
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-1629
 Alignment Scores:
 Pred. No.: 2,12e-21 Length: 1090
 Score: 453.00 Matches: 84
 Percent Similarity: 77.55% Conservative: 30
 Best Local Similarity: 57.14% Mismatches: 33
 Query Match: 8.59% Indels: 0


```
DB: 4 Gaps: 0
US-09-914-042-1 (1-1006) x US-09-270-767-1629 (1-1090)
Qy 2 ProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAlaPro 21
Db 443 CCATCCCTGATTGCGGTGTCGAGTTTGTGGAGGAGACGCGCTCCGATTACAGCTCGCCC 384
Qy 22 ThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIleGlu 41
Db 383 ACCACTCCACCTTCGCGCTCGAGGATGCCGAGTCCCGCACACACGATTGGCGCTCTTGGAA 324
Qy 42 GluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAlaIle 61
Db 323 GAGATTGGAGTTCCAGCGGAGGGTCTAACTAAGCTGAAAAAGCGGTCAAGGCTATC 264
Qy 62 AsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeuGluLys 81
Db 263 CACAACCTCAGGAAACACCCATGTGACATGAGATGTTTATGTCGCGAGCTCTCGAGCGA 204
Qy 82 PheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLysPhe 101
Db 203 CTGGCGCGCAAGGTTATTGACGAGGATGAGCGCGACATCGCGCGCGTTCCTTAAGTTC 144
Qy 102 SerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsn 121
Db 143 AGCGTTGTCACCAAGAGCTTAGCGCGCTGATGAAGACCTGATGCAAAACATCAACAAC 84
Qy 122 IleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGlyAsp 141
Db 83 ATTGTGATGTTTCGGTGGACTCAATGCTGAAGAGCGAGCTCGCGGCGTGAAGGGGAC 24
Qy 142 LeuLysLysProPheAspLys 148
Db 23 ATGAAGCGCCATTCGACAAG 3
RESULT 13
US-09-828-303-11
; Sequence 11, Application US/09828303
; Patent No. 6677504
; GENERAL INFORMATION:
; APPLICANT: COSTA E SILVA, OSWALDO DA
; APPLICANT: BOHNET, HANS J.
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, ROUYING
; TITLE OF INVENTION: TRANSCRIPTION FACTOR STRESS-RELATED PROTEINS AND
; FILE REFERENCE: 16313-0030
; CURRENT APPLICATION NUMBER: US/09/828,303
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/196,001
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2041
; TYPE: DNA
; ORGANISM: Physcomitrella patens
US-09-828-303-11
Alignment Scores:
Pred. No.: 2e-19 Length: 2041
Score: 429.00 Matches: 154
Percent Similarity: 39.30% Conservatives: 103
Best Local Similarity: 23.55% Mismatches: 217
Query Match: 8.14% Indels: 180
DB: 4 Gaps: 23
US-09-914-042-1 (1-1006) x US-09-828-303-11 (1-2041)
Qy 134 AspLeuLysGlyValIysGlyAspLeuLysProPheAspLysAlaTrpLysAspTyr 153
Db 14 GACTTGGGAATGCAAAATCACA-----TTTGAGCAGGCTCGATTCAATTG 61
Qy 154 GluThrLysIleThrLysIleGluLysGluLysGluHisAlaLysLeuHisGlyMet 173
Db 62 ATGACGACCTTACCAATAGTAGGAGCAAAAAAGAGTTTCGAG----- 103
Qy 174 IleArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGluArgPhe 193
Db 104 TTCCTTGAAGCGGTGAGTGTACA-----ATGGATGCACATCTCAGGTAC 148
Qy 194 PheGlnLeuGlnMetCysGluTyrLeuLysValAsnGluIleLysIleLysGly 213
Db 149 TTC-----AAGCAGGCG 160
DB: 4 Gaps: 23
US-09-914-042-1 (1-1006) x US-09-270-767-16911 (1-1090)
Qy 2 ProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAlaPro 21
Db 443 CCATCCCTGATTGCGGTGTCGAGTTTGTGGAGGAGACGCGCTCCGATTACAGCTCGCCC 384
Qy 22 ThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIleGlu 41
Db 383 ACCACTCCACCTTCGCGCTCGAGGATGCCGAGTCCCGCACACACGATTGGCGCTCTTGGAA 324
Qy 42 GluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAlaIle 61
Db 323 GAGATTGGAGTTCCAGCGGAGGGTCTAACTAAGCTGAAAAAGCGGTCAAGGCTATC 264
Qy 62 AsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGluLys 81
Db 263 CACAACCTCAGGAAACACCCATGTGACATGAGATGTTTATGTCGCGAGCTCTCGAGCGA 204
Qy 82 PheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLysPhe 101
Db 203 CTGGCGCGCAAGGTTATTGACGAGGATGAGCGCGACATCGCGCGCGTTCCTTAAGTTC 144
Qy 102 SerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsnAsn 121
Db 143 AGCGTTGTCACCAAGAGCTTAGCGCGCTGATGAAGACCTGATGCAAAACATCAACAAC 84
Qy 122 IleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValIysGlyAsp 141
Db 83 ATTGTGATGTTTCGGTGGACTCAATGCTGAAGAGCGAGCTCGCGGCGTGAAGGGGAC 24
Qy 142 LeuLysLysProPheAspLys 148
Db 23 ATGAAGCGCCATTCGACAAG 3
RESULT 12
US-09-270-767-16911/c
; Sequence 16911, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16911
; LENGTH: 1090
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16911
Alignment Scores:
Pred. No.: 2.12e-21 Length: 1090
Score: 453.00 Matches: 84
Percent Similarity: 77.55% Conservatives: 30
Best Local Similarity: 57.14% Mismatches: 33
Query Match: 8.59% Indels: 0
DB: 4 Gaps: 0
US-09-914-042-1 (1-1006) x US-09-270-767-16911 (1-1090)
Qy 2 ProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIlysAlaPro 21
Db 443 CCATCCCTGATTGCGGTGTCGAGTTTGTGGAGGAGACGCGCTCCGATTACAGCTCGCCC 384
Qy 22 ThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIleGlu 41
Db 383 ACCACTCCACCTTCGCGCTCGAGGATGCCGAGTCCCGCACACACGATTGGCGCTCTTGGAA 324
Qy 42 GluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAlaIle 61
```



```
Db 695 AGCAGCATGGAAGAGCGTTACGATTTTATCCGCGCCAAAGTATGTTGCCAAGCGCTATGTC 636
Qy 537 ArglylsYsHsAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaVallysthr 556
Db 635 ATGCGCAGCTGCTCGGATGACAAACGACTTACGGTGCAGCTTAGAGCAGCGTGTGCTCAAC 576
Qy 557 ArgAspIlePheGlyLeuLeuGlnAlaTyAlaAspGlyValAspLeuThrGluLysile 576
Db 575 GCCGACATGATGACTGCTCAGGTGTGGCGGAGGAGCGGATCTCAGCTGCTGCTG 516
Qy 577 ProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerVal 596
Db 515 CCC-----AGCTCTGATCGCGGAGAGCGGCCCTCCATCTGCGCGTGTGCGCGAG 465
Qy 597 AspArgThrSerLeuHisIleValAspPheLeuValGlnAsn----- 610
Db 464 ATGGGATCCAGCTGCACATAGTACATTTCTCATCCAAATATATGCCCCCTAAAGGCTT 405
Qy 611 -----SerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAla 624
Db 404 AACAAAGTACCAATCCGCGCGCTCTGTGGAT---GTGACGGGAAG---AATACGGCG 351
Qy 625 LeuHisTyCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLys 644
Db 350 TTGCATTGTGGCTCTGCAGATCGGAGGAGTGCATGAAGCTGCTGCTCGCTCTGGA 291
Qy 645 AlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeu 664
Db 290 GCCGACTATGACTCAAGAATCTACAGAAACAAACAGCTCTGGATATTGCCAAGAAATG 231
Qy 665 LysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHis 684
Db 230 GGACACATAGCTGCGAGGAGCTTATGATGTGCAATTAGCGGGAGAGCGCCCTTC 171
Qy 685 ValHisValGluTyTrpArgLeuLeuHisGluAsp-----LeuAspGluSerAsp 702
Db 170 GACCACATCAACACTGACTGGAATCTGCCAAACAGGAGCGGTCCACCGATTTCAGCGAC 111
Qy 703 AspAsp-----MetAspGluLysLeuGlnProSer 712
Db 110 GATGAACAGCATCATCGATGAGCGCTCCAGTTCTAGT 75
```

RESULT 15

```
Sequence 35, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: No. 6569662el Nucleic Acids and
FILE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
```

```
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 35
LENGTH: 2119
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (520)..(1614)
US-09-620-312D-35
Alignment Scores:
Pred. No.: 4e-12 Length: 2119
Score: 320.00 Matches: 108
Percent Similarity: 38.07% Conservative: 50
Best Local Similarity: 26.02% Mismatches: 159
Query Match: 6.07% Indels: 98
DB: 4 Gaps: 10
```

```
US-09-914-042-1 (1-1006) x US-09-620-312D-35 (1-2119)
Qy 310 GlySerLeuTyLysLysSer---AspGlyIleArgLysValTrpGlnLysArgLysCys 328
Db 409 GGGATCTCTCTAAAGCGGAGCGCAAGTCCCTGCAACAGGAGTGGAGAAGAAGTATGTG 468
Qy 329 SerVal---LysAsnGlyPheLeuThrIleSer----- 338
Db 469 ACGTCTGTGCACACGGGCTCTCACCTATCACCCAGCCTGCATGATTATCATGCAGAAC 528
Qy 339 ---HisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 529 ATCCACGGC-----AAGGAGATTGACCTGCTGCGGCAACACGCTGAAA 570
Qy 357 ----- 357
Db 571 GTGCGAGGAAGCGCTGCCCGAGCCACACCTGCCACAGCCCCCGGACCAGCCCCCGT 630
Qy 358 -----ThrAsnProGlu 361
Db 631 GCCAACGGGCTGCTCGTGGAGCGGAGTAACACACAGCTGGTGGTGGGGGCACAGGCGACAG 690
Qy 362 GluLysLysCysPheAspLeuIleSer---HisAspArgThrTyHisPheGlnAlaGlu 380
Db 691 GAGTCGTTTGAATTGTGGTGTCTCCCTCACTGGCAGACGCTGGCACTTCGAGGCTTCA 750
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluLysAlaLeu 400
Db 751 ACGGCGGAGGAGCGGAGCTGTGGGTTTCAGAGTGTGCAGGCCCCAGATCCTTGGCCAGCCTG 810
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 811 CAAGCTGCCAGCTGCCAAGGACAAGACTCGACTGGGGAACACAGACGCGAGCTCTGGCT 870
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 871 -----GTGACGGCGCTCCGACCGCTCCGCGGCAACAGACTTTTGTATCGACTGCGAT 921
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 922 GCACCAATCCAGACTGGGCCAGCCTGAACCTGGGTCGCCCTGTGATGTCATTGAGTGCTCA 981
Qy 461 GlyIleHisArgGluLeuGlyValHisTySerProMetGlnSerLeuThrLeuAspVal 480
Db 982 GGCATCCACCGACACTGGGGGCTCACCTGTCCCGGGTGGCGCTCCCTTGACTCGATGAC 1041
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1042 TGGCGCGCTGAGCTGCTGGCTGTCATGACTGCCATGGCAATGCTCCGCGCAACAGCGTC 1101
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1102 TGGGAGGGGGCTTTGGGTGGCTACTCC-----AAGCCAGGCGCTGTGCTCGACAGAG 1155
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 4, 2005, 08:27:30 ; Search time 1229 Seconds
(without alignments)
4845.614 Million cell updates/sec

Title: US-09-914-042-1

Perfect score: 5273

Sequence: 1 MPDQISVSEFVAETHDYKA.....DGDPRKGAPVSVFHIAD 1006

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool_h/US0914042/runat_04082005_090631_2104/app_query.fasta_1.1159
-DB=N-Geneseq -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MAIRIX=blowum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0914042 -CGN 1 1 615 @runat_04082005_090631_2104 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3874	73.5	5954	2	AAV59105
2	3417	64.8	4328	2	AAQ28758 Zebrafish
3	3155	59.8	4595	2	AAV59106 Zebrafish
4	3029	57.4	3775	10	ADD71214 Human int
5	3024.5	57.4	3946	10	ADD71207 Human int

6	3013.5	57.1	5330	2	AAV59103
7	3006	57.0	3974	12	ADJ51008
8	2954.5	56.0	4382	2	AAV59104
9	2852	54.1	5947	13	ADS34277
10	2556.5	48.5	4997	13	ADS10088
11	2555.5	48.5	5033	4	AAI59224
12	2455.5	46.8	4870	13	ADS34272
13	2410	45.7	5475	13	ADS34273
14	2089.5	39.6	3812	10	ADBG2827
15	2089.5	39.6	4050	10	ADKG1946
16	1803.5	34.2	3830	4	ABL04309
17	1801.5	34.2	3436	4	ABK43447
18	1801.5	34.2	3436	12	ADI53834
19	1759.5	33.4	3413	4	ABK43768
20	1759.5	33.4	3413	12	ADI54155
21	1583	30.0	6688	4	ABL04308
22	1307.5	24.8	3123	10	ADBG2160
23	802	15.2	1435	4	AA333079
24	766.5	14.5	544	13	ADQ79413
25	695	13.2	563	4	AA58478
26	611	11.6	2563	6	ABN59873
27	587	11.1	1224	8	ACA57215
28	581.5	11.0	5319	6	ABN83974
29	573.5	10.9	3926	10	ADD71213
30	571.5	10.8	6990	4	AAI58278
31	571.5	10.8	6990	5	ADQ98485
32	571.5	10.8	6990	9	ADB48245
33	570.5	10.8	2484	10	ABX70672
34	561	10.6	2737	6	ABK88008
35	558	10.6	5421	12	ADQ25213
36	552.5	10.5	6979	5	AA581429
37	534	10.1	2877	13	ADS09828
38	532.5	10.1	2505	6	AB567744
39	532	10.1	747	6	ABK71808
40	509.5	9.7	2494	6	ABK83644
41	509.5	9.7	2494	13	ADS34261
42	509.5	9.7	2498	13	ADS34262
43	509	9.7	2223	13	ADS34264
44	509	9.7	6480	6	ABK83626
45	493.5	9.4	7027	4	AAI60064

ALIGNMENTS

RESULT 1
AAV59105
ID AAV59105 standard; cDNA; 5954 BP.
XX
AC AAV59105;
XX
DT 17-OCT-2003 (revised)
DT 20-NOV-1998 (first entry)
XX
DE Zebrafish differentiation enhancing factor 2 gene.

ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;
C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;
adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;
nervous system.

OS Danio rerio.

XX Key Location/Qualifiers

XX CDS 433..3381

XX FT /*tag= a

XX FT /product= "DEF2 protein"

XX PN WO9836065-A1.

XX PD 20-AUG-1998.

XX PF 13-FEB-1998; 98WO-US002724.

XX

PR 14-FEB-1997; 97US-0038191P.
 XX (DAND) DANA FARBER CANCER INST INC.
 XX Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;
 XX WPI, 1998-467173/40.
 DR P-PSDB; AAW77288.
 XX New nucleic acid encoding differentiation enhancing factor - used
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating
 PT tumours and neurological disease.
 XX Claim 7; Fig 14; 203pp; English.
 XX The differentiation enhancing factors (DEF), comprise at least one each
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
 CC are mediators of SH3-domain dependent signalling and may be involved in
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,
 CC endocytosis or adhesion, migration, proliferation and differentiation of
 CC cells. Typical applications of DEF and agents that modulate interaction of
 CC between the protein and it's ligand, or of nucleic acid expressing them,
 CC are treatment of hyperplastic and neoplastic disease (a wide range of
 CC solid tumours and leukaemias) including metastases; for in vitro
 CC induction of differentiation of neural crest cells to neurons, glial
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in
 CC the nervous system (e.g. treatment of traumatic injury, stroke,
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral
 CC sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 5954 BP; 1731 A; 1382 C; 1372 G; 1469 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4,89e-162 Length: 5954
 Score: 3874.00 Matches: 755
 Percent Similarity: 83.12% Conservative: 92
 Best Local Similarity: 74.09% Mismatches: 122
 Query Match: 73.47% Indels: 50
 DB: 2 Gaps: 13
 US-09-914-042-1 (1-1006) x AAV59105 (1-5954)
 QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
 DB 433 ATGCTGACCAAGATTAACATGTCGGAGTTTGTACCGAGACAAATGAAGATTATAAATCG 492
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
 DB 493 CCCACCGCTCANAACCTTCAACACCAAGATGACTCACTGACGAAACACAGTATCCGCACTG 552
 QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
 DB 553 GAGGAGGCCCTGGATGTGACCGCAGTGTCTTTACAAAGATGAAGATCAGTTAAGGCT 612
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInTyrThrGlnAlaLeuGlu 80
 DB 613 ATTTACGGCTCGGGTCTGGCTCATGTGGAGATGAGGAGCATACACTCAAGCTCTGGAG 672
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
 DB 673 AAGTTGGAGAGAACTGTGTGTACAGAGATGACCCGACCTTGGGATCAGCTTCTCTGAAG 732
 QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
 DB 733 TTCTCCGCTTCCACCAAGAGCTCAGGCACCTCTTCAAGAACCTGTTCAGAACATGAAT 792
 QY 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
 DB 793 AATATCAVTTACCTTCCCATTTGACAGTCTGTGTAAGGGAGATCTGAAAGGGGTTAAAGGG 852
 QY 141 AspLeuLysLysProPheAspLysAlaIleLysAspTyrGluThrLysIleThrLysIle 160

DB 853 GATCTCAAGAAAGCCCTTCGATAAAGCTGGAAAGACTACGAGACTAAAGTCTCTAAATA 912
 QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
 DB 913 GAGAAGGAGAAAAAAGAGCAGCCCGCAGCAGCGAATGATCCGAGCGAGATCAGCGGA 972
 QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCysGlu 200
 DB 973 GCAGAGATAGCAGAAGAGATGGAAAAAGAGCGCGTTCTTCCAGCTTCAGATGTGTAG 1032
 QY 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
 DB 1033 TACCTCTCAAAAGTCAATGAATCAAGATCNAANAAGGTGTGACCTGTCTCAGATCTC 1092
 QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
 DB 1093 ATCAAAATACTTCCACGACAGTGCACCTCTTTTTCAGGATGGTCTCAAAAGCGGTGACAAC 1152
 QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
 DB 1153 CTCAAAACCTCAATAGAAAAACTCGCCACAGACTTGCATCGATCAAAACACAGGTACAGAT 1212
 QY 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
 DB 1213 GAAGAACGACAGACAGCTTACCAGTTACGGAGTGTGTNAANAAGTGTCTGCAAGTGGAG 1272
 QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
 DB 1273 CAGAAGGAGGACTCTCAGGTAGACAGAGCGCCACCTACAGTCTGCACACCGCAGCGGC 1332
 QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLeuArg 320
 DB 1333 AACAAAGAGCATGGGACTGAGCGCAGCGGCAACCTTTACAAAGAGAGTGAAGCGGTGCGG 1392
 QY 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
 DB 1393 AAGTGTGGCAGAGAGAGAGTGCACAGTAAAGAAATGGATATTTGACCATCTCACATGGG 1452
 QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
 DB 1453 ACGGCAAAACAGACCTCCGCGCAAACTCAATCTTCTCACTGTCTCAGGTGAAGACAAACCCA 1512
 QY 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
 DB 1513 GAGGAGAAGAAAAAGTTTGTACCTCATCTCACATGACAGAACATATCATTTTCCAGCGAGAA 1572
 QY 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
 DB 1573 GATGAGCCAGAGTGTCAANATATGGATCTCAGTGTGCGAGAACAGTAAAGAAGAGCGGCTC 1632
 QY 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
 DB 1633 AACACGCTTCAAGGGCGACAGCATGTTGGTGAATAAATCAATTTGTGCGAGAGCTCAC 1692
 QY 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
 DB 1693 AAGGCCATCTTGGGAGAGGTGAAGCGATGCGGGGAACGATGTCTGCTGCGACTGCCGT 1752
 QY 441 AlaProAspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
 DB 1753 GCTCCCGGCCCAATGGCTCTCCACCAACTCGGGCATCTCGACTGCATGCATGAGTTCG 1812
 QY 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
 DB 1813 GGGATCCACAGAGAGCTGGGCGTCCATTTACTTCCGGAATCCAGTCCCTCACACTCGACGTC 1872
 QY 481 LeuGlyThrSerGluLeuAlaIleAsnIleGlyAsnAlaGlyPheAsnGluIle 500
 DB 1873 CTCAGACCTCCGAGCTCTTGTCTGGCCAAAGAACGTGGGGAATGCTGGCTTCAATGAGATC 1932
 QY 501 MetGluCysLysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520


```
Db 1933 ATGGAGGCGCTGCTGACGCGCAGAGAGTGTATCAACCGAATCCAGCCAGTGCACATGCAG 1992
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgTyrAlaArgLysHis 540
Db 1993 GCGAGGAAGGACTTTATCATGCGCAAAATACACAGAGAAACGCTTCGCTCGTAAGAAGTGT 2052
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 2053 CCAGACGCACTGTGCAAGCTGCACACGCTGTGTGATGCTGTGAAGCCCGGACATTTTC 2112
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2113 TCTCTCATCCAGGTCTATGCTGAAGAGTGTGATCTGTAGGCCCATCTCTCTGCTAAT 2172
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2173 GGACATGAACAAGGTGAGACGGCTCTTCATCTGGCGGTGAGACTGGTGGACAGAACTTCC 2232
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2233 CTACATCATCTGAGCTTCTCCACCAAAACAGTTTAAACCTGGATAAGCAACCGCTAAA 2292
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2293 GGAACACAGCTCTGCATCTATCTGCTGCTGACGACACAGCGAGTGCTCAAACTGCTG 2352
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2353 CTCAGAGAAAGCCTCCATAGATATCGTATGAAGCTGGAGACACCCCGTTGGACATC 2412
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2413 GCCAGCGCATCAAAATCTGCAGTGTGAGGAACCTGCTGAACACGAGCTCTTGCAGGAAG 2472
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2473 TTCAATGCTCATGTCATGTGAGTATGAGTGGAGACTTTCAGCTGAGTCCAGAGACCTGGACGAG 2532
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2533 AGTGATGAAGATCTGATGAGAGTCGATCTCT-----CACCGCGGGATGAGCGGCC 2586
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2587 ATCAGCTGCTACACACCGCGGAGTAACTCCCTTCAGCTGAGTCCAGCGCAGCTGAGCGGA 2646
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2647 GACGCTCAGACCTGGTTAAAGACAGCAACGCG---TTGTGCCAAACCTGGTCAACAAT 2703
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
Db 2704 GAAACCTACGGGACCATTAATACACAGCTCACCC-----GTACGCGCTG 2748
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
Db 2749 TCCTCTTCTGCTCCACCTCTACACCCCGAAACCTA-----GTTACGCGCTGCTCTT 2802
Qy 801 AlaAsnThrLeu-----TrrLysThrAsnSerValSerValAspGly 815
Db 2803 CGAGGACTGACTCAAGGATCTCCCGGCTGGAGCCTGCTCTCTGGATCTGAGCGGC--- 2859
Qy 816 SerArgGlnArgSerSerAspProProAlaValHisProProLeuProLeuArg 835
Db 2860 ---AGACAGAGATCTCTCTGACCTCCCAACATGCATCTCTCTGCGCCCTCCCTTACGG 2916
Qy 836 ValThrSerThrAsnProLeuThrProThr---ProProProProValAlaLysThrPro 854
Db 2917 GTCACTTCACCTCTTCTAATGCCAGCGGTGCTGCTCTCTCTCTCTGCTAAAGCTACT 2976
Qy 855 SerValMetGluAlaLeuSerGlnProSerLysPro-----AlaProGlyIleSer 872
Db 2977 GGTATGATGGAGACCATGAATATGCAACCCCAACCCGACAGCGGGCGCTCTCTGGACAGAAC 3036
```

```
Qy 873 GlnIleArgProProProLeuProGlnProProSerArgLeuProGlnLysLysPro 892
Db 3037 ATCAACCGG----- 3045
Qy 893 AlaProGlyThrAsp-----LysSerThrProLeuThrAsnLysGlyGlnPro 908
Db 3046 GCTCAAGTGGCGGCAAAAACTTCAGCAAAAGCACACTGATGCGCTCC----- 3093
Qy 909 ArgGlyProValAspLeuSerAlaThrGluAlaLeu---GlyProLeuSerAsnAlaMet 927
Db 3094 ---GGATCCATCGAGACGACCTAAAGAGTCCCGAGGAGGCCCAACACCACTGGT 3150
Qy 928 ValLeuGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLys 947
Db 3151 CAAATCTGCTGCGCACCCACATGCCAGGAAA-----ACGTATTTGAAGCCGAAG 3201
Qy 948 ArgValLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGlu 967
Db 3202 CGTGTGAAGGCCATGTATACTGTGTGCCGATATATCCAGACGAGCTGACCTTCTCTGAG 3261
Qy 968 GlyAspValIleIleValAspGlyGluGluAspGlnGluTyrTrpIleGlyHisIleAsp 987
Db 3262 GGAGAGCTTATCGTGTGATGGAGAGGAGGACGAGAGTGTGCTGGCCACATTGAG 3321
Qy 988 GlyAspProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3322 GGAGAGCCAATGAGAAGAGGAGCGTTCTCTGTACAGTTGTACAGTTTCAATTATGGAC 3378

RESULT 2
ID AAQ28758 standard; DNA; 4328 BP.
XX AC AAQ28758;
XX XX
DT 25-MAR-2003 (revised)
DT 25-FEB-1993 (first entry)
XX DE
XX KW
XX OS Homo sapiens.
XX PN WO9215602-A1.
XX PD 17-SEP-1992.
XX PF 28-FEB-1992; 92MO-US001624.
XX PR 28-FEB-1991; 91US-00662216.
XX PA (DAND ) DANA FARBER CANCER INST INC.
XX PI Sager R;
XX DR WPI; 1992-331663/40.
XX PT Diagnosis and treatment of cancer - using candidate tumour suppressor
XX PT genes or the corresp. antibodies.
XX PS Claim 30; Page 38-41; 54pp; English.

An adaptation of the subtractive hybridization technique was used which
utilises a biotinylation-based subtraction procedure instead of
hydroxyapatite as previously used. In this procedure, a single strand
phagemid cDNA library from normal cell polyA+ mRNA is hybridized with
excess biotinylated tumor polyA+ mRNA, and the resulting double stranded
sequences are removed by binding to streptavidin. The remaining single-
stranded phagemid cDNAs are converted to double-stranded form and used to
transform bacterial host cells. The resulting subtracted cDNA library is
differentially screened with total cDNA from normal and tumor cells. This
method produced some 20 additional cloned cDNAs. Also found by this
```

CC method were several genes which, on the basis of the partial DNA
 CC sequences appear to be novel sequences not previously entered into
 CC GENBANK. The portion of the cDNAs so sequenced represents part of the
 CC coding region and/or part of the 3' untranslated region of each cDNA (see
 CC Q28749-58). (Updated on 25-MAR-2003 to correct PN field.) (Updated on 25-
 CC MAR-2003 to correct PA field.)

XX Sequence 4328 BP; 1236 A; 970 C; 912 G; 1210 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5,436-142 Length: 4328
 Score: 3417.00 Matches: 648
 Percent Similarity: 99.39% Conservative: 1
 Best Local Similarity: 99.23% Mismatches: 3
 Query Match: 64.80% Indels: 1
 DB: 2 Gaps: 0

US-09-914-042-1 (1-1006) x AQAQ28758 (1-4328)

QY 355 GlnValysThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThr 374
 DB 3 CAGGTGAAGACCAACCCCTGAGGAGAAAGAGTCTTTGACCTTATTTCACATGACAGAACT 62
 QY 375 TyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsn 394
 DB 63 TACCACCTTCAAGCTGAAGATGAACAGGAATGTCAATATGATGCTGTGCTGCAAAAT 122
 QY 395 SerLysGluGluAlaLeuAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsn 414
 DB 123 AGCAAGAAGAAGCTTTAAACAATGCATTTAAGGGGGATGACATATCTGGAGAAATTAAC 182
 QY 415 IleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAsp 434
 DB 183 ATCGTCCAAAGAACTGACAAAGAGAGATCATCTCAGAAAGTGACAGAGATGACGGGCAATGAC 242
 QY 435 ValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeu 454
 DB 243 GTCTGCTGTGACTGTGGGGGCCAGATCTTACATGCTGCTTCCACCAACTGGGCATCTCG 302
 QY 455 ThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMetGln 474
 DB 303 ACCTGCTCAGTGTTCGGGAATCCACGAGAGCTGGGGGTTCATTATTCAGGATGACG 362
 QY 475 SerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyVal 494
 DB 363 TCCCTGACCTTAGATGTGGAACTCTGAGCTGCTGCTCGCCCAAGAAATATTGGGAAT 422
 QY 495 AlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsn 514
 DB 423 GCAGGCTTAAATGAGATCATGGAATGTTGCCCTACCCAGCTGAGGACTCAGTCAACCCCAAC 482
 QY 515 ProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg 534
 DB 483 CCAGCAGCGACATGAATGCAAGAAGAGCTACATCACAGGCAAGTACATCGAGAGGAGA 542
 QY 535 TyrAlaArgLysLysHisAlaAspAsnAlaLysLysHisSerLeuCysGluAlaVal 554
 DB 543 TACGCAAGAGAAGACACGCGGATTAACCGCGGAAGCTTCAAGTCTTTGGAGGCCGCTC 602
 QY 555 LysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGlu 574
 DB 603 AAACGAGAGATATTTTGGATGCTCCNAGCTTATGCTGATGCTGTGATCTTACGGA 662
 QY 575 LysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArg 594
 DB 663 AAAATCCCACTGGCCCAAGGACATCAGCGGATGAACGGGCCCTCCACCTTGCAGTCAGA 722
 QY 595 SerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerClyAsnLeu 614
 DB 723 TCCGTGGATGCAACCTCTCTTCACATTTAGACTTTTATGTTCAAGACAGTGGGAACTG 782
 QY 615 AspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAla 634

DB 783 GATAAACAGACAGGAGAAAGGAGCAGCACAGCCCTGCTACTGCTGCCTGACCGACAATGCC 842
 QY 635 GluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGly 654
 DB 843 GAGTGCCTCAAGTTGCTCTCTCGGGGGAAGCCCTCCATCGAGATAGCAAAATGAGTCAGGA 902
 QY 655 GluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThr 674
 DB 903 GAGACTCCGTGGACATTGCCAAGCGCTCAGACACGAGCACTGTGAGGAGCTGCTGACC 962
 QY 675 GlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTyrArgLeuLeu 694
 DB 963 CAAGCCTTATCTGAAGATTTAATTTCTCAGCTTCAAGTGAATATGAATGCGGACTACTC 1022
 QY 695 HisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGluAsn 714
 DB 1023 CACGAAGACCTGGATGAAGTGAATGATGATGATGAGAAATTCGACCCAGTCCCAAC 1082
 QY 715 ArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsn 734
 DB 1083 CGGCGGGAAGACCCGCCCATCAGCTTCTACAGCTGGGCTCCAAACCAAGCTTCAGTCTAAC 1142
 QY 735 AlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMet 754
 DB 1143 GCTGTATCTTTGGCCAGAGATGCTGCAACCTTGCCAAAGGACAGCAGAGGGCTTTCATG 1202
 QY 755 ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerProProAla 774
 DB 1203 CCCAGCATCTTGCAAGATGAGACTTACGAGGCCCTCTCTGAGTGGCAGCCCACTCCCGCC 1262
 QY 775 GlnProAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794
 DB 1263 CAGCTCGAGCCCCCAGACCAACAGCGGCCCCCCCTCTCTCCAGGAAATGTGGCAA 1322
 QY 794 sValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGly 814
 DB 1323 AGTTTCAGACAGCCCTCTCTGCTAACACCCCTGTGGAAGACAAACTCTCTAGTGTGGACGG 1382
 QY 814 YGlySerArgGlnArgSerSerSerAspProProAlaValHisProProLeuProProLe 834
 DB 1383 TGGAAAGCGGCGAGCGATCTTCTGTCAGATCCGCGAGCTGTCCATCCACGCTGCCCTCT 1442
 QY 834 uArgValThrSerThrAsnProLeuThrProThrProProProValAlaLysThrPr 854
 DB 1443 TCGGTGACATCTACCAATCCCTTGACCCCGCCCGCCCGCCCGTGGCCAGAGCGCC 1502
 QY 854 oSerValMetGluAlaLeuSerGlnProSerLysProAlaProProGlyLysSerGlnI 874
 DB 1503 CAGCGTAATGGAAGCCTTGAGCCAGCGAGCAAGCCTGCCCGCTGCGGATCTCACAGAT 1562
 QY 874 eArgProProProLeuProProGlnProProSerArgLeuProGlnLysLysProAlaPr 894
 DB 1563 CAGGCCCCCACTCTGCCCCACAGCGCCCGCCCGCTCCCGCAGAAAGCGCTGCGGCC 1622
 QY 894 oGlyThrAspLysSerThrProLeuThrAsnLysGlyGlnProArgGlyProValAspLe 914
 DB 1623 AGGGGCTGACAAATGTCACCCCACTGACCAACAAAGGCGAACCCGAGAGACCTGTGGATCT 1682
 QY 914 uSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaPr 934
 DB 1683 CTCTGCAACGGAAGCTCTGGGTCTCTGTCTCAATGCTATGCTCTGCGAGCCCTGCAACC 1742
 QY 934 oMetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAs 954
 DB 1743 CATGCTTAGGAAGTCGAGGCAACCAAGTTGAAGCCTTAAGCGGGTGAAGAGCGCTCTATAA 1802
 QY 954 nCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAs 974
 DB 1803 CTGTGTGGCTGACAAACCCCGATGAGCTCACCTTCTCCGAGGGGATGTGATCATCTGGA 1862
 QY 974 pGlyGluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGly 994
 DB 1863 CGGGAGGAGGAGCAGGAGTGGTGGATTGGCCACATTTGATGATGATGATGATGATGATGATG 1922


```
Qy 341 ThrAlaAsnArgProAlaLysLeuLeuLeuThrCysGlnValLysThrAsnPro 360
|||
Db 1320 ACGCCCATGACCGCCAGCAAACTGAACCTCTTAACCTGCCAAGTGAAGCGTAACCCA 1379
|||
Qy 361 GluGluLysCysPheAspLeuLeuSerHisAspArgThrTyrHisPheGlnAlaGlu 380
:::
Db 1380 GATGAGAAAAAATGCTTTGATCTCATATCATGACAGACAGACGATCATCTCCAGACTGAG 1439
|||
Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
|||
Db 1440 GATGAGCGAGATGTCAGTATGGGTTCTCTGTTCTCCAGAACAGTAAGAGAGCGGCTG 1499
|||
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
|||
Db 1500 AACATGCTTTAAAGACATCAGATGAGGGAGGAGAAATAATCATTTGTCGAGAGCTCACT 1559
|||
Qy 421 LysGluIleLeuSerGluValGlnMetThrGlyAsnAspValCysCysAspCysGly 440
|||
Db 1560 AAGGCCATCGTGGGGGAAGTGAAGAAATGAGCGCAATGACGTGTGCTGTGACTGTGGA 1619
|||
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGluIleLeuThrCysIleGluCysSer 460
|||
Db 1620 GCTTCCAAATCCAAACATGCTCTCCAAACCTGGGTGTGTTGATTGCAATTGAATGCTCT 1679
|||
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
|||
Db 1680 GGGATCCATCGGAAATGGGGTCCACTACTCCGAATACAGTCTCTGACACTGGACCTC 1739
|||
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
|||
Db 1740 TTAGGCACATCTGAACACTATGCTGCTAACAGTGTGGGAAATGACGATTCATTAATGAAATC 1799
|||
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
|||
Db 1800 ATGGAAGCAAACTGCTCTCTCAGAG---ATCCAAAACCTTACCTTCTAGTGACATGCAG 1856
|||
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
|||
Db 1857 GTACGAAAGACTTCATCACGCCAAATACACAGAGAGCGTTTCGCTCAGAGAGAGTAT 1916
|||
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
|||
Db 1917 GCAGATAACGCGAGCTCGAGTGCATGCTGATGATGATGAGTGAAGTCTCGGACATCTTC 1976
|||
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
|||
Db 1977 TCCCTGATCCAGGCTATGCTGAAGGACTGGACCTGATGGAGACCAATTAATCAGCCTAAC 2036
|||
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
|||
Db 2037 CAACATGAACCCAGCGGAGACATCATCTATCTTGGCGTACGATGCTGGACCGAAACTCC 2096
|||
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
|||
Db 2097 CTCCTATATTGGACTTTCTTGTACAGAACAGTGGCAATTTAGACAAAGCAGACGCCAAA 2156
|||
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
|||
Db 2157 GGAAGCACAGCGCTACATATTGCTGCTGATGATGATGATGATGATGATGATGATGATG 2216
|||
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
|||
Db 2217 CTGCGGGGGAAGCATCTGTACGATTAATGATGCTGGAGACTGCTCTGATTTG 2276
|||
Qy 661 AlatyArgLeuLysHisGluHisCysGluGluLeuLeuLeuThrGlnAlaLeuSerGlyArg 680
|||
Db 2277 GCGCAGCGTCTCAAACTCCAAATGCGAGGAGCTGCTGACTCAGCGCGCAGACGGGGAAG 2336
|||
Qy 681 PheAsnSerHisValHisValGluTyrGluThrArgLeuLeuHisGluAspLeuAspGlu 700
|||
Db 2337 TTCATGTCCATGTGCATGTGGAATATGACTGGCGTCTGCATTAATGAGGATCTGACCGAG 2396
|||
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgGluAspArgPro 720
```

```
Db 2397 AGCGAAGATGAGATGGAGGACAAG-----CCCATTCCTCATCGGCGTGAGGAGCTCCA 2450
|||
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
|||
Db 2451 ATAGCTGTATAGTTCAGGCGAGTGGCCCATGATGATGCCCAACATCAGGCGCTCTGGCTCGG 2510
|||
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
|||
Db 2511 GACGTGCCAATGTGGTCAATATAAGCAGAGGCTTTTATTCGAGCATGATGATGATGATG 2570
|||
Qy 761 GluThrTyrGlyAlaLeuLeuSerGlySerProProAlaGlnProAlaAlaProSer 780
|||
Db 2571 GAGACTTACGGCACCATGCTCGATCCCACTCTCCACCACTG-----GGTTTACCAGGA 2624
|||
Qy 781 ThrThrSerAlaProProLeuProProArgAsnValGlyLysValGlnThrAlaSerSer 800
|||
Db 2625 GTACCTGGCATTCCTCTTTTACCCCTCGGCCCTTGGGAAGGGA----- 2669
|||
Qy 801 AlaAsnThrLeuTyrLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSer 820
|||
Db 2670 -----TGG-----AGTCCACCAATGGAGAACATCGGTAGACAGAGTCA 2708
|||
Qy 821 SerSerAspProProAlaValHisProProLeuProProLeuArgValThrSerThrAsn 840
|||
Db 2709 TGTTTCAGATCTCGCA-----AACCTCAAACTCTCGAACAAATAAATACTCTGTGTAT 2759
|||
Qy 841 ProLeuThrProThrProProProProValAlaLysThrProSerValMetGluAlaLeu 860
|||
Db 2760 GTTCTGCTCTCTGCTCTCTCCACCTCTCTGCA----- 2792
|||
Qy 861 SerGlnProSerLysProAlaProProGlyLysSerGlnIleArgProProProLeuPro 880
|||
Db 2793 -----CCCAAGACACTCCACTCCAGATCCAAAGCCAGCTCTTCTCTCCAGCAGCC 2846
|||
Qy 881 ProGlnProProSerArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThr 900
|||
Db 2847 ACGGCTCT-----CCTGCACCA-----TCGCA 2870
|||
Qy 901 ProLeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeu 920
|||
Db 2871 CCGCTCTTATTCACCTGCTCTCTCTCAGGCGACGCGCT----- 2909
|||
Qy 921 GlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetProArgLysSerGln 940
|||
Db 2910 -----GTAGTGGCCCTCGACCCAGTATGCCCATTCCTCGTCA 2945
|||
Qy 941 AlaThrLysLeuLysProLysArgValLysAla 951
|||
Db 2946 CTGACTGATGTCAAAAGTCTGCTCTCTAAAGCC 2978
|||
RESULT 4
ADD71214
ID ADD71214 standard; cDNA; 3775 BP.
XX
AC ADD71214;
XX
XX 15-JAN-2004 (first entry)
DT
DE Human intracellular signalling molecule INTS1G-51 cDNA SEQ ID NO:103.
KW human; intracellular signalling molecule; INTS1G; cytosolic;
KW antiarteriosclerotic; antidiabetic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
KW antiinflammatory; thymimetic; gene therapy;
KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;
KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;
KW immune disorder; inflammatory disorder; AIDS; allergy;
KW developmental disorder; Hypothyroidism; Cushing's syndrome;
KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.
OS Homo sapiens.
XX
```

PN WO2003039348-A2.
 XX 15-MAY-2003.
 XX 07-NOV-2002; 2002WO-US036151.
 XX 09-NOV-2001; 2001US-0344472P.
 PR 30-NOV-2001; 2001US-0334558P.
 PR 14-DEC-2001; 2001US-0340296P.
 PR 21-DEC-2001; 2001US-0343557P.
 PR 18-JAN-2002; 2002US-0350420P.
 PR 25-JAN-2002; 2002US-0351927P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Kable AB, Swarnakar A, Corvad AE, Hafalia AJA, Duggan BM;
 PI Warren BA, Emerling BM, Ison CH, Nguyen DB, Lindquist EA, Lee EA;
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li JX;
 PI Marquis JP, Gietzen KJ, Baughn MR, Borowsky ML, Yao MG, Chawla NK;
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;
 PI Lee SY, Tran UK, Elliott VS, Sprague WW, Tang YT, Zebajadian Y;
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;
 PI Zheng W;
 XX WPI; 2003-441441/41.
 DR P-PSDB; ADD71162.
 XX New human intracellular signaling molecules (INTSIG)), useful for
 PT diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant INTSIG expression e.g. cancer, diabetes, epilepsy, or
 PT infections.
 XX Claim 5; SEQ ID NO 103; 363pp; English.
 XX The present invention describes human intracellular signalling molecules
 CC designated INTSIG-1 to INTSIG-52. The INTSIG polynucleotides and proteins
 CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
 CC antiinflammatory and thymomimetic activities, and can be used in gene
 CC therapy. The INTSIG polynucleotides and proteins can be used in the
 CC diagnosis, treatment and prevention of diseases or conditions associated
 CC with the decreased expression or overexpression of INTSIG, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),
 CC immune/inflammatory (e.g. AIDS, allergies), developmental (e.g.
 CC hypothyroidism, Cushing's syndrome), reproductive and vesicle-trafficking
 CC disorders, or infections. They are also useful in assessing the effects
 CC of exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of INTSIG. The INTSIG or its fragments are useful in screening
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,
 CC or in altering the expression of the target polynucleotide and compounds
 CC that specifically bind to or modulate the activity of the polypeptide.
 CC The present sequence encodes human INTSIG-51 from the present invention.
 XX
 SQ Sequence 3775 BP; 1123 A; 989 C; 873 G; 790 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 6, 47e-125 Length: 3775
 Score: 3029.00 Matches: 619
 Percent Similarity: 69.78% Conservative: 129
 Best Local Similarity: 57.74% Mismatches: 238
 Query Match: 57.44% Indels: 86
 DB: 10 Gaps: 19
 US-09-914-042-1 (1-1006) x ADD71214 (1-3775)
 QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
 DB 318 ATGCGCGACCATCTCTGCTCGAGTTCATCGCGGACACCCGAGGACTACAACTCG 377
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 40
 DB 378 CCCACCAAGTCCAGCTTACCACCGGGTGCACAACTGCGAGAACACCGCTCACTGCTG 437

QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
 DB GAGGAGGCTCTAGACCAAGATAGAACGCCCTTCAGAAAGTGAAGAGCTGTCTAAAGCA 497
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
 DB ATATATAATCTGGTCAAGATCATGTACAAAATGAAGAAAACATATGCACAACTTCTTAT 557
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
 DB AAGTTGGAGTAAATTTTAAAGTCGACACACCCCGACCTTGGCACCCTTCTTCTCAAG 617
 QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuLysGlnAsnMetAsn 120
 DB TTTTCTACTCTTACAAAGAACTGTCCACACTGCTGAAAAAATCTGCTCCAGGTTTGAGC 677
 QY 121 AsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLysGly 140
 DB CACATGTGATCTTTCACCTTGGATTTCTTTGTTAAAGGAGACCTTAAGGGAGTCAAGGA 737
 QY 141 AspLeuLysLysProPheAspLysAlaLysAspTyrGluThrLysLysLysLysLys 160
 DB GATCTCAAGAACCCATTTGACAAAGCTTGGAAAGATTATGACAAAGTTTACAAATTT 797
 QY 161 GluLysGluLysLysGluHisAlaLysLysLysGlyMetIleArgThrGluLysSerGly 180
 DB GAGAAAGAGAAAGAGAGACGACGCAAAACACATGGGATGATCCGCACAGAGATAACAGGA 857
 QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnGlnGlnMetCysGlu 200
 DB GCTGAGATTGGGAGAAATGGAGAAAGGAGGCGCTCTTTTTCAGCTCCAAATGTGTGAA 917
 QY 201 TyrLeuLeuLysValAsnGluLysLysLysLysLysLysLysLysLysLysLysLys 220
 DB TATCTCATTAAGTTTAAATCAAGACCAAAAGGGGTGGATCTGCTGCAGATCTT 977
 QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
 DB ATAAAGTATTACCATGCACAGTGCATTTCTTCAAGATGGCTTCAAAACAGCTGATAAG 1037
 QY 241 LeuLysProSerIleGluThrIleuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
 DB TTGAACAGTAGTACATTTGAAATTCGCTCTGCTGATTTATATATATATATATATATAT 1097
 QY 261 GluGluArgArgGlnLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 280
 DB GAAGAAAGAAACAGCTAACTGCACTCCGAGACTTAAATAATCTCTTCTCACTGGAT 1157
 QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
 DB CAGAAAGAAAGATTTCTCAGAGCGGCAAGGA--GGATACAGCATGCATCAGCTCCAGGCG 1214
 QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLysArg 320
 DB AATAGGAATATGGCAGTGAAGAGGGGTACCTGTCTTAAAGAAAGTGAAGGAGTCCCG 1274
 QY 321 LysValTyrGlnLysArgLysCysSerValLysValAsnGlyPheLeuThrLysHisGly 340
 DB AAGTATGCGAGAGGAGGAAGTTCAGTCAAGAAATGGGATTTCTGACCATCTCAGATGCC 1334
 QY 341 ThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
 DB ACATCTAACAGGCAACCAAGTGAACCTTCTCACCCTGCCAAGTAAACCTAATGCC 1394
 QY 361 GluLysLysLysCysPheAspLeuIleSerHisAspArgThrThrHisPheGlnAlaGlu 380
 DB GAAGCAAAAAATCTTTTGCCTGATATACATATAGAACATATACATCTTTCAGGAGAA 1454
 QY 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluLysLysLys 400
 DB GATGAGCAGGATTATGTAGCATGGATATCATGATTATGACAAATAGCAAGAGAGGCCCTA 1514

```
Qy 401 AsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAenAenIleValGlnGluLeuThr 420
Db 1515 ACCATGGCCCTCCGTGGAGACAGAGTCGGGAGAGAACACCGTG--GAAGACCTGACA 1571
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAenAspValCysCysAspCysGly 440
Db 1572 AAAGCCATTATTAGGATGTCACGGCTCCACGGGATGACATTTGCTGCGATTGTGGC 1631
Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1632 TCATCAGAACCCACCTGCGCTTTCAACCAACTGGGTATTTTGACCTGATAGAATGTTCT 1691
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1692 GGCATCCATAGGAAATGGGGTTTCATATTTCTCGCATTCAGTCTTTGGAACTAGACAAA 1751
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1752 TTAGGAACCTCTGAACCTCTGCTGGCCCAAGATGTAGGAACAATAGTTTAAATGATATT 1811
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1812 ATGGAAGCAAAATTTACCCAGC---CCCTCACCAAAACCCACCCCTTCAAGTGATAGACT 1868
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1869 GTACGAAAGAATATATCACTGCAAAAGTATGTAGATCATAGTGTTCGAAGAACACTGT 1928
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 1929 TCAACTTCATCAGCTAAATAAATGAATTGCTTGAGGCCATCAATCCAGGGATTTACTT 1988
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 1989 GCACATAATCAAGTCTATGCGAAGGGGTAGAGCTAATGGAACCACTG---CTGGAACCT 2045
Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2046 GGGCAGGAGCTTGGGAGACAGCCCTTCACCTTGGCCGTCCGAACCTGCAGATCAGACATCT 2105
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2106 CTCATTATGGTTGACTCTCTGTACAAACCTGTGGGAACCTGGATAAGCAGAGCCCGCTG 2165
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2166 GGAACACAGACTTCTACACTACTGTAGTATGTACAGTAAACCTGAGTGTTCGAAGCTTTG 2225
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2226 CTCAGAGCAAGCCACTGTGATATAGTTAACAGGCTGGAGAACTGCCCTAGACATA 2285
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2286 GCAAAAGACATAAAAGCTACCCAGGTGGAAGATCTGCTTCCAGGCTAAATCTGGAAG 2345
Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
Db 2346 TTCAATCACACGTCGACGTAGAAATAGTGTGAATCTTCGACAGGAGGATAGATGAG 2405
Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAenArgGluAspArgPro 720
Db 2406 AGCGATGATGATCTGGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCACCCAGACT 2465
Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
Db 2466 CAGAGCTTCTGCCACTCTCTCCAGCATCTCCCCCAGGAC-----AAGCTGGCACTG 2516
Qy 741 AspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
Db 2517 CCAGGATTCAGCACTCCCAAGGGACAAACAGCGGCTC----- 2552
Qy 761 GluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGlnProAla 777
```

```
Db 2553 ---TCCTATGGAGCCTTCCACCAACAGATCTTCGTTTCCCAAGCACAGACTCCGCCACA 2609
Qy 778 AlaProSerThrThrSerAlaProLeuProProArgAenValGlyLysValGlnThr 797
Db 2610 TCACCA---ACCACGAGGCTCCCTCTCTAGGAACGCGGGAAGGTAACGAT 2666
Qy 798 AlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlySerArg 817
Db 2667 GGGGGTCCATCTCTTCAAGTAAGACTACAAC-----AAGTTTGAGGGACTATCCAG 2720
Qy 818 GlnArgSerSerAspPro-----ProAlaValHisProProLeuPro 832
Db 2721 CAGTCGAGACCAAGTCTTGCCAAAGACTGCCCTTGGCCCAAGAGTCTTCTCTAAACTACT 2780
Qy 833 -----ProLeuArgValThrSerThrAsnProLeu----- 842
Db 2781 CAGAAAGTGGCACTAAGGAAAAACAGATCATCTCTCCCTAGACAAAGCCACCATCCCGCC 2840
Qy 843 -----ThrProThrProProProPro----- 849
Db 2841 GAAATCTTTCAGAAATCATCACAGTTGGCAGAGTTGCCACAAAGCCACCACTTGGAGAC 2900
Qy 850 -----ValAlaLysThrProSerValMetGluAlaLeuSerGln 862
Db 2901 CTGCCCCCAAGCCCAAGCAACTGGCCCCCAAGCCCAATTTGGAGATTTTCCCGCCTAAG 2960
Qy 863 ProSerLys---ProAlaProProGlyIleSerGlnIleArgProProPro----- 878
Db 2961 CCAGGAGAACTGCCCCCAACCAACACAGCTGGGGGACCTGCCCAACCAACCACTCTCA 3020
Qy 879 ---LeuProProGlnProProSerArg---LeuProGlnLys----- 890
Db 3021 GACTTACTCTCCAAACCAACAGATGAAGACCTGCCCCCCCAACCAACCAAGCTGGAGACCTG 3080
Qy 891 -----LysProAlaProGlyThr 896
Db 3081 CTAGCAAAATCCCAGACTGGAGATGTCTCACCCAAGGCTCAGCAACCTCTCTGAGGTGACA 3140
Qy 897 AspLysSerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeu 914
Db 3141 CTGAAGTCAACACCCATTTGGATCTATCCCAAAATGTGCAGTCCAGAGACGCCATCCAAAAG 3200
Qy 915 SerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProAlaPro 934
Db 3201 CAAGCATCTGAAGACTCCAGACCTCAGGCTACTCTGCCA---GAGACGCCGTACCA 3257
Qy 935 MetProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsn 954
Db 3258 CTGCCCCAGAAAAATCAATACGGGAAAAATAAAGTGAGGGGAGTGAAGACCATTTATGAC 3317
Qy 955 CysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleValAsp 974
Db 3318 TGCCAGGCAGACACACGATGACGAGCTCACATTCATCGAGGGGAGAAGTGAATATCGTCACA 3377
Qy 975 GlyGluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGly 994
Db 3378 GGGGAAGAGGACCAAGAGTGGTGGATTGGCCACATCGAAGGACAGCTTGAAGAGAGAGGGG 3437
Qy 995 AlaPheProValSerPheValHisPheIleAlaAsp 1006
Db 3438 GTCTTCCAGTGTCTTTGTTTCATATCTCTGTCTGAC 3473

RESULT 5
ADD71207
ID ADD71207 standard; cDNA; 3946 BP.
XX
AC ADD71207;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human intracellular signalling molecule INTSIG-44 cDNA SEQ ID NO:96.
XX
```


KW human; intracellular signalling molecule; INTS1G; cytoskeletal;
 KW antiarteriosclerotic; antidiabetic; anticonvulsant; neurotropic;
 KW neuroprotective; cerebroprotective; anti-HIV; antiallergic;
 KW antiinflammatory; thyroidemic; gene therapy;
 KW cell proliferative disorder; cancer; atherosclerosis; endocrine disorder;
 KW diabetes; neurological disorder; epilepsy; Huntington's disease; stroke;
 KW immune disorder; inflammatory disorder; AIDS; allergy;
 KW developmental disorder; Hypothyroidism; Cushing's syndrome;
 KW reproductive disorder; vesicle-trafficking disorder; infection; gene; ss.

XX Homo sapiens.
 XX WO2003039348-A2.
 XX 15-MAY-2003.
 XX 07-NOV-2002; 2002WO-US036151.
 XX 09-NOV-2001; 2001US-0344472P.
 PR 30-NOV-2001; 2001US-0334558P.
 PR 14-DEC-2001; 2001US-0340296P.
 PR 21-DEC-2001; 2001US-0343557P.
 PR 18-JAN-2002; 2002US-0350420P.
 PR 25-JAN-2002; 2002US-0351927P.
 XX (INCY-1) INCYTE GENOMICS INC.
 PA Kable AB, Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM;
 PI Warren BA, Emerling BM, Leon CH, Nguyen DB, Lindquist EA;
 PI Yue H, Yue H, Forsythe IJ, Ramkumar J, Griffin JA, Li UX, Lee EA;
 PI Marquis JP, Gierzen KJ, Baugman MR, Borowsky ML, Yao MG, Chawla NK;
 PI Lehr-Mason PM, Lal PG, Gururajan R, Khare R, Batra S, Becha SD;
 PI Lee SY, Tran UK, Elliott VS, Sprague WW, Tang YT, Zebartjadian Y;
 PI Jiang X, Jackson AA, Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A;
 PI Zheng W;
 XX WPI; 2003-441441/41.
 DR P-PSDB; ADD71155.
 XX New human intracellular signaling molecules (INTS1G), useful for
 PT diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant INTS1G expression e.g. cancer, diabetes, epilepsy, or
 PT infections.
 XX Claim 5; SEQ ID NO 96; 363pp; English.
 PS The present invention describes human intracellular signalling molecules
 CC designated INTS1G-1 to INTS1G-52. The INTS1G polynucleotides and proteins
 CC have cytostatic, antiarteriosclerotic, antidiabetic, anticonvulsant,
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic,
 CC antiinflammatory and thymomimetic activities, and can be used in gene
 CC therapy. The INTS1G polynucleotides and proteins can be used in the
 CC diagnosis, treatment and prevention of diseases or conditions associated
 CC with the decreased expression or overexpression of INTS1G, such as cell
 CC proliferative (e.g. cancer, atherosclerosis), endocrine (e.g. diabetes),
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),
 CC immune/inflammatory (e.g. AIDS, allergies), developmental (e.g.
 CC hypothyroidism, Cushing's syndrome), reproductive and vesicle-trafficking
 CC disorders, or infections. They are also useful in assessing the effects
 CC of exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of INTS1G. The INTS1G or its fragments are useful in screening
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,
 CC or in altering the expression of the target polynucleotide and compounds
 CC that specifically bind to or modulate the activity of the polypeptide.
 CC The present sequence encodes human INTS1G-44 from the present invention.
 XX Sequence 3946 BP; 1165 A; 1064 C; 902 G; 815 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,078-124 Length: 3946
 Score: 3024.50 Matches: 627
 Percent Similarity: 66.84% Conservative: 127
 Best Local Similarity: 55.59% Mismatches: 233

Query Match:

57.36%

Indels:

141

DB:

10

Gaps:

21

US-09-914-042-1 (1-1006) x ADD71207 (1-3946)

QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIysAla 20
 DB 318 ATGCCGGACCAAGATCTCTGTCGGAGTTTCATCGCCGAGACACCCGAGACTCAACTCG 377
 QY 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaLeu 40
 DB 378 CCCACACGTCAGCTTCCACACGGGGTGCACAACTGCAGGAAACACCGTCAGCTGCTG 437
 QY 41 GluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerValIysAla 60
 DB 438 GAGGAGGCTTAGACCAAGATAGAACAGCCCTTCAGAAAGTGAAGAAGTCTCTAAAGCA 497
 QY 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGluTyrThrGlnAlaLeuGlu 80
 DB 498 ATATATATATCTGTCGTCAGATCATGTACAAATGAAGAAACTATGCAACAGTTCTTGAT 557
 QY 81 LysPheGlyGlyAsnCysValCysArgAspProAspLeuGlySerAlaPheLeuLys 100
 DB 558 AAGTTGGAGTATTTTAACTCGACACACCCCGACCTTGGCACCCCGTTTGTCAAG 617
 QY 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
 DB 618 TTTTCTACTTTACAAAGGAACTGTCCACACTGCTGTAATAATCTGCTCCAGGTTTGAGC 677
 QY 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValIysGly 140
 DB 678 CACAATGTCCTTCCACCTTGGATCTTTTGTAAAGGAGAGACTAAAGGAGTCAAGGA 737
 QY 141 AspLeuLysLysPheProPheAspLysAlaTrpLysAspTyrGluThrLysLeuVal 160
 DB 738 GATCTCAAGAGCCATTTGACAAAGCTTGAAGAAGATTATGACAAAGTTTCAAAAT 797
 QY 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluLysSerGly 180
 DB 798 GAGAAAGAGAAAGAGAGAGACGCAACAAACATCGGATGATCCGCACAGAGATAACAGGA 857
 QY 181 AlaGluIleAlaGluGluMetGluLysGluArgArgPheGluLeuGlnMetCysGlu 200
 DB 858 GCTGAGATTGCGAAGAAATGAGAAAGGAAAGGCGCTCTTTTCAGCTCCAAATGTGTAA 917
 QY 201 TyrLeuLeuLysValAsnGluLysLeuLysLysGlyValAspLeuLeuGlnLeu 220
 DB 918 TATCTCATTAAGTTTAATGAATCAAGACCAAAAGGGTGTGGATCTGCTGAGAAATCTT 977
 QY 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLysAlaValGluSer 240
 DB 978 ATAAAGTATTACCATGCACAGTGCATTTCTTCAAGATGGCTTCGAAACACAGCTATAG 1037
 QY 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
 DB 1038 TTGAAACAGTACATTTGAAATACTGCTGCTGATTTATATATATATAATAAAACAGACCCAGAT 1097
 QY 261 GluGluArgArgGlnLeuIleGlnLeuArgPheLeuLysSerAlaLeuGlnValGlu 280
 DB 1098 GAAGAAAGAAAGAACAGCTAACTGCACTCCGAGACTTAATAATAATCTCTCTCACTGGAT 1157
 QY 281 GlnLysGluAspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
 DB 1158 CAGAAAGAAAGATCTCAGACGCGGCAAGGA---GGATACAGCATGCATCAGCTCCAGGCG 1214
 QY 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyLeuArg 320
 DB 1215 AATAAGGAATATGCGAGTGAAGAGGGGTACCTGCTTAAAGAAAGTGAAGCGGATCCGG 1274
 QY 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
 DB 1275 AAAGTATGCGAGAGGAGGAGGTGTTTCACTCAAGATGGGATTTCTGACCATCTCAGATGCC 1334

Db 3441 ATCAATACGGGGAAAAATAAGTGAGCGGAGCTGAAGACCATTATATGACTGCCAGGCAGAC 3500
 Qy 959 AsnProAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAsp 978
 Db 3501 AACGATCAGCAGAGCTCACATTCATCGAGGAGAGTGAATTCGTACAGGGGAAGAGCAG 3560
 Qy 979 GlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgIysGlyAlaPheProVal 998
 Db 3561 CAGGAGTGGTGGATTGGCCACATCGAAGGACAGCCTGAAGGAGAGGGGGCTTTCCAGTG 3620
 Qy 999 SerPheValHisPheIleAlaAsp 1006
 Db 3621 TCCCTTTGTTCAATCCTGTCTGAC 3644
 RESULT 6
 ID AAV59103
 ID AAV59103 standard; cDNA; 5330 BP.
 AC AAV59103;
 XX
 DT 20-NOV-1998 (first entry)
 XX
 DE Bovine differentiation enhancing factor 1 gene.
 DE
 KW ss; bovine; differentiation enhancing factor; ankyrin repeat; C2 domain;
 KW SH3 consensus binding sequence; pleckstrin homology domain; adipogenesis;
 KW neurogenesis; hyperplastic disease; neoplastic disease; nervous system.
 XX
 OS Bos sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 209..3598
 FT /*tag= a
 FT /*product= "DEF1 protein"
 XX
 PN W09836065-A1.
 XX
 PD 20-AUG-1998. 98WO-US002724.
 XX
 PF 13-FEB-1998; 98WO-US002724.
 XX
 PR 14-FEB-1997; 97US-0038191P.
 XX
 PP (DAND) DANA FARBER CANCER INST INC.
 XX
 PI Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;
 XX
 DR WPI; 1998-467173/40.
 DR P-PSDB; AAW77286.
 XX
 PT New nucleic acid encoding differentiation enhancing factor - used
 PT particularly to regulate adipogenesis and neurogenesis, e.g. for treating
 PT tumours and neurological disease.
 XX
 PS Claim 10; Fig 2; 203pp; English.
 XX
 CC The differentiation enhancing factors (DEF), comprise at least one each
 CC of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
 CC domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
 CC are mediators of SH3-domain dependent signalling and may be involved in
 CC cellular gene expression, cytoskeletal architecture, protein trafficking,
 CC endocytosis or adhesion, migration, proliferation and differentiation of
 CC cells. Typical applications of DEF and agents that modulate interaction of
 CC between the protein and it's ligand, or of nucleic acid expressing them,
 CC are treatment of hyperplastic and neoplastic disease (a wide range of
 CC solid tumours and leukaemias), including metastases; for in vitro
 CC induction of differentiation of neural crest cells to neurons, glial
 CC cells etc.; for increasing neuron survival, and inducing cell repair, in
 CC the nervous system (e.g. treatment of traumatic injury, stroke,
 CC Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral
 CC sclerosis, multiple sclerosis etc.)
 XX
 SQ Sequence 5330 BP; 1416 A; 1502 C; 1276 G; 1136 T; 0 U; 0 Other;

1166	CTGCAGGGCCAAACAGGAATATCGGCAGCGAGAGAGAGGGCTACTCTGCTGAAGAAGAGTGAC	1225	Db
318	GlyIleArgLysValTrpGlnLysArgLysCysSerValIysAsnGlyPheLeuThrIle	337	Qy
1226	GGGATCCGGAAGATGTGGCAGAGAGAAGTGTCTCGTCAAGAACGGGATCTCTGACCATC	1285	Db
338	SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys	357	Qy
1286	TCCACGCCACGTCCTCAACACAGACGACGCAAGCTGAACCTTCTCACTGCCAGGTGAAG	1345	Db
358	ThrAsnProGluLysLysCysPheAspLeuIleSerHisAspArgThrTyHisPhe	377	Qy
1346	CCGAATCCCGAGGACAAGAAGTCTTTGACCTGATATCACATAACAGGACGTATCACTTT	1405	Db
378	GlnAlaGluAspGluGlnLysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu	397	Qy
1406	CAGGCCGAGAATGAGCAGGATATGTAGCTGTGATCTCGGTGCTGCAAAACAGCAAGAAG	1465	Db
398	GluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsnAsnIleValGln	417	Qy
1466	GAGGCCCTCACCATGGCTTCGCGGGGGAACAGAGTGTCTGGGAGAGCAGCCCTG--GAG	1522	Db
418	GluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCys	437	Qy
1523	GAGCTGACGAAGGCCATCATCGAGAGCTGCAGCGCTCCCGGGCAACGACGTCTCTCTGC	1582	Db
438	AspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIle	457	Qy
1583	GACTCGCGCTCGGCAGAACCCACCTGCTGTCCACCAACTTTGGGCATCTTGACTGTATA	1642	Db
458	GluCysSerGlyIleHisArgGluLeuGlyValHisTySerProMetGlnSerLeuThr	477	Qy
1643	GAATGTTCCGCATCCATAGAGAAATGGGGTTTCATATTTCTCGCATCCAGTCTTTGGAA	1702	Db
478	LeuAspValLeuGlyThrSerGluLeuLeuLeuAlaLysAsnIleGlyAsnAlaGlyPhe	497	Qy
1703	CTAGACAAATTTAGGAACCTTCGAACCTCTGCTGCCAAGAATGTAGGAAACAATAGTTT	1762	Db
498	AsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySer	517	Qy
1763	AATGATATTATGGAAGCAAAATTTACCAGT---CCCTCACAAACCCACCCCTTCAAGT	1819	Db
518	AspMetAsnAlaArgLysAspTyIleThrAlaLysTyIleGluArgArgTyAlaArg	537	Qy
1820	GATATGACTGTACGGAAGGAATATATCACTGCAAAGTATGTAGATCATAGTTTTCACGG	1879	Db
538	LysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArg	557	Qy
1880	AAGACCTGTTTCATCGTCTCAGCTAAACTGAAACGAATTCGCTTGGGCCATCAAAATCCAG	1939	Db
558	AspIlePheGlyLeuLeuGlnAlaTyAlaAspGlyValAspLeuThrGluLysIlePro	577	Qy
1940	GAITTTACTTGCACTAAATTCAGTCTATGACAGAGGGGTGAGCTAATGGAAACCGCTG--	1996	Db
578	LeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAsp	597	Qy
1997	CTGGAACCCGACAGAGACTTGGGGAGACAGCCCTTCATCTTGAGTCCGAACCGCAGAC	2056	Db
598	ArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGln	617	Qy
2057	CAGACATCTCTCCATTTGGTGGACTTCTCTGTACAAACCTGTGGAACTTAGATAAGCAG	2116	Db
618	ThrGlyLysGlySerThrAlaLeuHisTyCysCysLeuThrAspAsnAlaGluCysLeu	637	Qy
2117	ACGGCCCTGGGGAACACGGCCCTGCATCTACTGTAGTATGTACAGTAACACAGAGTGTG	2176	Db
638	LysLeuLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrPro	657	Qy
2177	AAGCTGTCTCAGGACGACGCCACTGTGGACGTCGTCTTAATCAGGCTGGAGAGACGCC	2236	Db
658	LeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeu	677	Qy
2237	CTGACATAGCAAGAGACTGAAGCACCCTCAGTGTGAAGACCTGCTTTCCCAAGCTAAA	2296	Db

Qy	678	SerGlyArgPheIasnSerHisValHisValGluTyrGluTirPArgIleuLeuHisGluAsp	697
Db	2297	TCTGGAAAGTTCAATCCACACGCTCCACGTTGGAAATATGAGTGGTAATCTTCGACAGSAGGAG	2356
Qy	698	LeuAspGluSerAspAspMetAspGluLeuLeuGlnProSerGluAenArgArgGlu	717
Db	2357	ATGATGAGAGCGATGACGACCTTGGATGACAAACGAGCCCATCAAGAGGAGCGCTCC	2416
Qy	718	AspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAenAlaValSer	737
Db	2417	CCCGACCGCAGAGCTTCTGCCACTCTCCAGCATCTCCCCCAGGAC-----AAG	2467
Qy	738	LeuAlaArgAspAlaalaAenLeuAlaLysGluLysGlnArgAlaPheMetProSerIle	757
Db	2468	CTTCACTCGCGGGCTTCACGACCGCAAGGGACAAGCAACGACTC-----	2512
Qy	758	LeuGlnAenGluThrTyrGlyAla-----LeuLeuSerGlySerPro	771
Db	2513	-----TCCATACGGCGCTTCACCAACCCAGACTCTTCGCTCCCAACGACACAGAC	2560
Qy	772	ProProAlaGlnProAlaIaProSerThrSerAlaProProLeuProProArgAsn	791
Db	2561	TCACCCACGTCACCGATCGCA-----GAGCGCGCCCGCTGCTCCCGACAAAC	2608
Qy	792	ValGlyLys-----ValGlnThrAlaSer	799
Db	2609	GCCACGAAAGTCCACTGGCCCCACTTCACACTCCCTCTAAGCACCCAGACTCTAGT	2668
Qy	800	SerAlaAenThrLeuTrpLysThrAsnSerValSerValaspGlyGlySerArgGlnArg	819
Db	2669	GGCAGCTCCACCTGTCCAAAGACGGTCTCTCTCCCCACACCCGGA---CACAAAGA	2725
Qy	820	SerSerSerAspProProAlaValHisProProLeuProProLeuArg-----ValThr	837
Db	2726	ACCTGTCTGACCTCTCCAGCCCACTACCTCACGGGCCCCCAACAAAGGCGCAGTCTCT	2785
Qy	838	SerThrAsnProLeuThrProThrProProProValAlaLysThrProSerValMet	857
Db	2786	TGGGTAAACACGCTGGGTCCCTCATCGTCC-----AGTAAGACCCACGAAACAGTTC	2836
Qy	858	GluAlaLeuSerGln-----	862
Db	2837	GAGGCGCTGTCCCGACGAGTCGAGCACCGGTTCTGCAAAAGACTGCATTTGTCCAAGATT	2896
Qy	863	-----ProSerLysProAla-----	867
Db	2897	CTTCCTAAACTACTCAGAAAGTGGGACTTAAGGAAACACAGACACCCATCATCTCTCC	2956
Qy	868	-----ProProGlyIle-----SerGlnIle-----	874
Db	2957	CTCGACAAAGCCAAACGTCCCACTGAGATCTTCCAGAAAGTCGTCCCAAGTTGACAGAGTTA	3016
Qy	875	-----ArgProProPro-----LeuProGlnPro-----	883
Db	3017	CCGCAAGAGCGGCACCCCGGGGACCTGCCCCCCGAAGCCCAAGGACCTGGCTCCCAACCC	3076
Qy	884	-----ProSerArgLeuProGlnLys-----	890
Db	3077	CCCATTGGAGACTTACCACCTAAGCCAGCGGAGCTGCCCCCGAAGCCACAGCTGGGCGAC	3136
Qy	890	-----	890
Db	3137	CTGCCCCCAGCCCCAGCTCGCAGACTTGCCCCCAGCCCCAGGTGAAGACCTGCGCT	3196
Qy	891	---LysProAlaProGlyThrAspLysThrProLeuThrAenLysGlyGlnProArg	909
Db	3197	CCCAAGCCACAACTGGGGGAGCTGTGGCAAAACCCCAAGACGGGAGACGCTCGCCCAAG	3256
Qy	910	Gly-----ProValAspLeuSer-----	915
Db	3257	GCCCAGCACCCCTGGAGCTCACCCCCAAGTCAACCCGCGGACCTGTCCCCGAAAGCTC	3316

Db 2565 -----TCTATGGAGCCTTCCACCAACGAGTCTTCGTTTCCACAGACGAC 2612
Qy 775 GlnProAlaProSerThrThrSerAlaProProLeuProProArgAsnValGlyLys 794
Db 2613 TCGCCCATCATCAACA---ACCAGGAGGCTCCCTCTGCCCCCTAGGAACGCGGAAA 2669
Qy 795 -----ValGlnThrAlaSerSerAlaAsn 802
Db 2670 GGTCAACTGGCCCACTTCAACATCCCTCTAAGCACCCAGACCTCTAGTGGCAGCTCC 2729
Qy 803 ThrLeuTpyLysThrAsnSerValSerValAspGlyGlySerArgGlnArgSerSer 822
Db 2730 ACCCTATCCAGAGAGGCTCTCTCCACACACCCGGA---CAACAGAGAACCTTATCC 2786
Qy 823 AspProAlaValHisProProLeuProLeuArg----- 835
Db 2787 GACCTCCAGCCCACTACCTCATGGGCCCCCAACAAAGGCGAGTCTCTTGGGGTAAC 2846
Qy 835 ----- 835
Db 2847 GATGGGGTCCATCTCTTCAAGTAAGACTACAAACAGTTTGGGGACTATCCAGCAG 2906
Qy 836 -----ValThrSerThrAsnPro---LeuThrProThrProProPro 848
Db 2907 TCGAGCACAGTCTTGCAAGACTGCTCTTGGCCCAAGAGTCTTCTCTAAACTACCTCAG 2966
Qy 849 ProValAla-----LysThr-----ProSer 855
Db 2967 AAGTGGGCACTAAGGAAACAGATCATCTCTCCCTAGACAAAGCCACCATCCGCGCGAA 3026
Qy 856 ValMetGluAlaLeuSerGln-----ProSerLysProAlaPro----- 868
Db 3027 ATCTTTCAGAAATCATCACAGTGGCAGAGTTGCCCAAAAGCCACCACTGGAGACCTG 3086
Qy 869 -----ProGlyIleSerGlnIleArgProProPro 878
Db 3087 CCCCCAAAGCCACAGAACTGGCCCCCAAGCCCCCAATTTGGAGATTTGCCGCCCTAAGCCA 3146
Qy 879 -----LeuProProGlnProPro---SerArg 886
Db 3147 GGAGNACTGCCCCCAACACAGCTGGGGGACCTGCCACCAACCACTCTCAGAC 3206
Qy 887 LeuProGlnLys----- 890
Db 3207 TTACTCTCCAAACACACAGATGAAGGACCTGCCCCCAACACACAGCTGGGAGACCTGCTA 3266
Qy 891 -----LysProAlaProGlyThrAsp 897
Db 3267 GCAAAATCCCAAGCTGGAGATGTTCTCACCCAGGCTCAGCAACCTCTGAGGTCACTG 3326
Qy 898 LysSerThrPro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSer 915
Db 3327 AGTCACACCATTTGGATCTATCCCAATGTGCAGTCCAGAGCGCCATCCAAAGCAA 3386
Qy 916 AlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLeuLeuGlnProProAlaProMet 935
Db 3387 GCATCTGAAGACTCCCAACAGCTCAGCGCTACTCTGCCA---GAGACGCGCTACCACTG 3443
Qy 936 ProArgLysSerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuThrAsnCys 955
Db 3444 CCGAGAAATCAATACCGGGGAAATAAAGTGGCGGAGTGAAGACCATTTATGACTGC 3503
Qy 956 ValAlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGly 975
Db 3504 CAGGAGAACACGATGACGAGCTCATCTATCGAGGGAGAGTATTCGTCTACAGG 3563
Qy 976 GluGluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAla 995
Db 3564 GAAGAGGACAGGAGTGGTGGATTGGCCACATCGAAGGACGACCTCGAAGGAGGGGTC 3623
Qy 996 PheProValSerPheValHisPheIleAlaAsp 1006
Db 3624 TTTCCAGTGTCTTTGTTTCATATCTCTGTCTGAC 3656

RESULT 8

AAV59104
ID AAV59104 standard; cDNA; 4382 BP.
XX
AC AAV59104;
XX 17-OCT-2003 (revised)
DT 20-NOV-1998 (first entry)
XX Zebrafish differentiation enhancing factor 1 gene.
XX ss; Zebrafish; differentiation enhancing factor; ankyrin repeat;
C2 domain; SH3 consensus binding sequence; pleckstrin homology domain;
KW adipogenesis; neurogenesis; hyperplastic disease; neoplastic disease;
nervous system.
XX Danio rerio.
XX
FH Key Location/Qualifiers
FT CDS 351..3806
FT /*tag= a
FT /product= "DEF1 protein"
XX
FN WO9836065-A1.
XX
PD 20-AUG-1998.
XX
PF 13-FEB-1998; 98WO-US002724.
XX
PR 14-FEB-1997; 97US-0038191P.
XX
PA (DAND) DANA FARBER CANCER INST INC.
XX
PI Thomas RM, King FJ, Harris DF, Hu E, Spiegelman B, Chan J;
XX WPI; 1998-467173/40.
DR P-PSDB; AAW77287.
XX
PT New nucleic acid encoding differentiation enhancing factor - used
particularly to regulate adipogenesis and neurogenesis, e.g. for treating
tumours and neurological disease.
XX
PS Claim 7; Fig 13; 203pp; English.
XX
CC The differentiation enhancing factors (DEF), comprise at least one each
of SH3 consensus binding sequence, ankyrin repeat, pleckstrin homology
domain and C2 domain. The DEF induce adipogenesis or neurogenesis; they
are mediators of SH3-domain dependent signalling and may be involved in
cellular gene expression, cytoskeletal architecture, protein trafficking,
endocytosis or adhesion, migration, proliferation and differentiation of
cells. Typical applications of DEF and agents that modulate interaction
between the protein and it's ligand, or of nucleic acid expressing them,
are treatment of hyperplastic and neoplastic disease (a wide range of
solid tumours and leukaemias), including metastases; for in vitro
induction of differentiation of neural crest cells to neurons, glial
cells etc.; for increasing neuron survival, and inducing cell repair, in
the nervous system (e.g. treatment of traumatic injury, stroke, in
Alzheimer's, Parkinson's or Huntington's diseases, amyotrophic lateral
sclerosis, multiple sclerosis etc.). (Updated on 17-OCT-2003 to
standardise OS field)
XX
SQ Sequence 4382 BP; 1222 A; 1132 C; 1098 G; 930 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1,46e-121 Length: 4382
Score: 2954.50 Matches: 622
Percent Similarity: 65.04% Conservative: 126
Best Local Similarity: 54.09% Mismatches: 239
Query Match: 56.03% Indels: 163
DB: 24 Gaps: 24
US-09-914-042-1 (1-1006) x AAV59104 (1-4382)

Qy 1 MetProaspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 411 ATGCCGGATCAGATCTCCGTCGTCAGATTCTCTCGAGAGCAGCGAGGATTCACAAATTC 470
Qy 21 ProThrAlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAlaIle 40
Db 471 CCCAGCCTCGAGCTTCACCAACCGCTCGAGAGCTCGCGGAACACCGTCAATGTTCTG 530
Qy 41 GluGluAlaLeuAppValAspArgMetValLeuTyrLysMetLysLysSerValLysAla 60
Db 531 GAAGAGGCTTTCGATCAGACCGAATCTCTTACAGAAGGTCAAGAAATCTGTCAAAGCA 590
Qy 61 IleAsnSerSerGlyLeuAlaHisValGluAsnGluGlnTyrThrGlnAlaLeuGlu 80
Db 591 ATCTACAACTCGGGTCAAGACATGTGCAGAAATGAAGAAATATTGGCAGGCATCGGAC 650
Qy 81 LysPheGlyGlyAsnCysValCysArgAspProaspLeuGlySerAlaPheLeuLys 100
Db 651 AAGTTTGCAGCAACTTCATCAGCCGAGATAACTCTGTATCTGGGAACAGCCTTCATCAAG 710
Qy 101 PheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMetAsn 120
Db 711 TTTTCTGGACTTATCAAAGAGCTGCTGCTCTCTCTCAAAGAACTGCTCCAGAGCCTCAGC 770
Qy 121 AsnIleIleSerPheProLeuAspSerLeuLysGlyAspLeuLysGlyValLysGly 140
Db 771 CACAACGTCATCTTCACCCTGGACTCTCTGCTCAAAGGAGATCTTAAAGGGAGTGAAGGG 830
Qy 141 AspLeuLysLysProPheAspLysAlaTyrLysAspTyrGluThrLysIleThrLysIle 160
Db 831 GACCTTAAAGCCCTTCGACAAAGCCCTGGAAGACTATGAACCAAGTTTCAAAAGATC 890
Qy 161 GluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGly 180
Db 891 GAGAAGGAGAAGAGAGACATGCCAAGCAGCAGCGCATGATCCGCACAGAAATCACCGGC 950
Qy 181 AlaGluIleAlaGluMetGluLysGluArgArgPheGlnLeuGlnMetCysGlu 200
Db 951 GCAGAGATTGCAGAAAGATGAGAAAGGAGCGGAGGATCTTTACAGTCGATGTGTGAG 1010
Qy 201 TyrLeuLeuLysValAsnGluIleLysIleLysLysGlyValAspLeuLeuGlnAsnLeu 220
Db 1011 TACCTGATCAAAGTCAATGAGATTAAAGCAAGAGGGAGTGGATCTCTCCAGATCTC 1070
Qy 221 IleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSer 240
Db 1071 ATCAAGTATTATCATGCACAGTGCATTTCTTCCAGGATGGCTTGAAAACTGCTGACAAG 1130
Qy 241 LeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAsp 260
Db 1131 TTGAAGCAGTATATTGAAAAATTTAGCAGCTGATCTTTATATATATAAACAAGACTCAGGAT 1190
Qy 261 GluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnValGlu 280
Db 1191 GAGGAGAAAAACAGCTCACAGCTCTCAGAGACCTCATCAATCTTCTTACAGCTGGAC 1250
Qy 281 GlnLysGluAppSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
Db 1251 CAGAAGGAGGATCTCAGAGTAAGCAGAGC---GGGTACAGCATGCACAGCTGCAGGCG 1307
Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
Db 1308 AATAAGGAGTTTGGCAGTGAGAAGGGCTATCTCTTCAAAGAAAGATGATGGGATCCGT 1367
Qy 321 LysValTyrGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
Db 1368 AAGTGTGGCAGAGAGGAGAGTGTCTCAGTGAANAATGGCATCCTCACCATCTCTCATGCC 1427
Qy 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAspPro 360
Db 1428 ACATCAAACAGCAGCGGTGAGACTGAATCTGCTGACTGCAGCGGTAAACCCAGTGA 1487

Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
Db 1488 GAGATTAAGAAGTCTCTTTGACCTCATCTCTCATTAATCGACATATCATTTCCAGGCAGAG 1547
Qy 381 AspGluGlnGluCysGlnIleTyrMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
Db 1548 GACGAACAGGAGTTTGTGTATGATCGGTGCTGACTAATAGTAGAAGGAGGAGCTCTG 1607
Qy 401 AsnAsnAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGluLeuThr 420
Db 1608 AACATGTCATTTCTGGGGAGCAGAGTCTGGAGATGACAGTTTG---GAGACTTGACC 1664
Qy 421 LysGluIleIleSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
Db 1665 AAACCATCATCGAGGAGTGTCTCGCATTCCTCGAAGCAGAGTCTGCTGTGACTGTGGG 1724
Qy 441 AlaProaspProThrTyrLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
Db 1725 GTTCAGAGCCCAAAATGGTTATCCACTAACCTCGGCATCTCTGAGGTGCATCGAGTGTTC 1784
Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
Db 1785 GGAATCCACAGGGAATGGGAGTCCATATTTCGGCGCATCCAATCCATGAGCTTGCACAAA 1844
Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
Db 1845 CTTCGAACTCTGAATCTCTGCTGCTTAGNACGTGGCAACAGTAGTTTCAACGAATA 1904
Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
Db 1905 TTAGAAGGGAATCTGCCGAGT---CCTTACCAAGCCAGCGCCATCAAGTGACATGACC 1961
Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
Db 1962 GAGAGGAAGGAGTACATCAATCGGAAGTACGTGAGCAGCAGGTTCTGCTCGCGCAACGCC 2021
Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
Db 2022 ACTACAGCCACAGCCAGCAGCGGCGACTTGTACAGGCGGTGAGNACCGAGAGACTTGATG 2081
Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
Db 2082 GCTCTCTTACGCTCTATGCAGATGGAGTGAGCTAATGGATCTCTTCCCAAGAACGA--- 2138
Qy 581 GlyHisGluProaspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
Db 2139 GGACAGACCGCGGAGAGACAGCTCTGACCTTTCTGTCGAGCATCAGACAGACTTCC 2198
Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
Db 2199 CTGCACCTGTGGAGTCTTCTGTCCAAAACAGTGGGACTCTAGACAGACACGAGAGT 2258
Qy 621 GlySerThrAlaLeuHisTyrCysCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
Db 2259 GGAACCGTCTCTCTTACTGCTGCACATATGAGAAGCCAGAGTGTCTCAAACTGCTG 2318
Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
Db 2319 CTCAGGGGAAAACCGTCTATTGACCTGTATTCAAAACCGGGAGACAGCATTTGGATATC 2378
Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
Db 2379 GCCAGACATCGAANAATGTACAGTGTGAAGAGTACTGTGTGGAGCGAGCCGGGAGG 2438
Qy 681 PheAsnSerHisValHisValGluTyrGluTyrArgLeuLeuHisGluAspLeuAspGlu 700
Db 2439 TTTAATCTCATGTGCATGTGAGTATGAGTGAATCTCGCGCTGGAGGAGATGTATGATGAG 2498
Qy 701 SerAspAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
Db 2499 AGTACCGATGACCTGGATGACAAAGCCTAGTCCAGTGAAGAGGAGCGTCTCTCGTCTCT 2558
Qy 721 IleSerPheTyrGlnLeuLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740


```

Db 2559 CAGAGCTTC-----TGTCATTGTCAGCGTGTCT-----CCTCAG 2594
Qy 741 AspAlaAlaAsnLeuAala-----LysGluLysGlnArg----- 751
Db 2595 GAGAAAGTTAAACCTCGCGGGTATCTAGGACACAGGGACAGAGACTGCTCTATGGA 2654
Qy 752 AlaPheMet----- 754
Db 2655 GCCTTTGCCAACCCCGCTTACAGCACCTCCACCGAAACCCCTGCTATCTCCAGTGTGCAGAG 2714
Qy 755 ---ProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGlySerPro----- 771
Db 2715 GGACCCACCATAGCCAGCAGAACCCCTGCAAAAGCTCCGCTCTGTGGCGCGCCACCTCT 2774
Qy 771 ----- 771
Db 2775 CTGCGCTGGGATCTCAATCGAGTCAGGAGCGAGCTCCACTTTGCTTAAGNAGAGAGCT 2834
Qy 772 ProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro----- 788
Db 2835 CCTCTCCACCTCCCGGACACAAAGGACCCCACTCAGATCCCGCCAGTCCCGTACTGCAG 2894
Qy 789 ---ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThr 807
Db 2895 GGTCCGACAGAAAGGAAGTAGTCCACACTCTCTGCAATTCGACATCCCGGCC 2954
Qy 808 AsnSerValSerValAspGlyGlySerArgGlnArgSerSerSerAsp----- 823
Db 2955 AAC-----AAGTTGAGGGATCCAGCAGCAGCAGCAAGCACTACGCTCTATGNAACAACAAA 3008
Qy 824 ---ProProAlaValHisProProLeu----- 831
Db 3009 GCAACATTTGGCCACGAGTTCTTCCCAAACTACCTCAAAAGTGGCACTACGAAAGATT 3068
Qy 831 ----- 831
Db 3069 GACAAATCCACCTCCCATCAGTGACAAAGTCTGTGCTGTGATGCTTTCAGAAACCCCA 3128
Qy 832 -----ProProLeuArgValThrSerThrAsnProLeuThr-----ProThr 845
Db 3129 CAGGCGCAGGATGCACCTCCACAGAGCCTCAGATACA-----ATAACAGACCCACT 3182
Qy 846 ProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 3183 GAACCTCCACCT-----AAAATTCACAGGTGCGAGAACGATCCCGCTGTGTGATGTC 3236
Qy 866 ProAlaProProGlyIleSerGlnIleArgProProPro-----LeuProPro 881
Db 3237 CCGCAGAAACCGCACATCTCAGACCTCTCCCAACCGCAACTATCAGATCTTCCCGCC 3296
Qy 882 GlnPro-----ProSerArgLeuProGlnLysLysProAlaProGlyThr 896
Db 3297 AAACCCCAATTTGCGATTACCAACCAAAA---CCTCAGCTTTCGACCTGCCCGCCGAG 3353
Qy 897 AspLysSerThrProLeuThrAsnLysGlyGln-----ProArgGlyProVal 912
Db 3354 CCTCAGCTTAAGGATCTTCCCGCTTAAGCGGAGATGATGCTGCCATCCAAACCGGCC 3413
Qy 913 AspLeuSerAlaThrGluAla----- 919
Db 3414 GTGTGTTCTGCTGTGAGGCCACACAGAGCGAGTCAACGCGAGGAGAAACCAAGTCCGAG 3473
Qy 920 -----LeuGlyProLeu----- 923
Db 3474 CCCAGCTGACGGAGACACAGTCATTCACGACGAGGAGAGCTCTACCCCGCAGCGCC 3533
Qy 924 -----SerAsnAlaMetValLeuGlnProProAlaProMetProArg 937
Db 3534 AGCGAGGACACCAATGGAGCCCGCCGAGGAGCTTGGAAATCCCAATGCCATGCCAGCC 3593
Qy 938 Lys-----SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysVal 956

```

```

Db 3594 AAAATTAACACAGTAGCAAGAACAAACGAGCGGTGTGAACCACTATGATGTCGAG 3653
Qy 957 AlaAspAsnProAspGluLeuThrPheSerGluGlyAspValIleIleValAspGlyGlu 976
Db 3654 GCAGACAATGACGATGAGCTGACTTTTGTGGAGGCGAGGTTATAATTTGTACACAGAGAG 3713
Qy 977 GluAspGlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe 996
Db 3714 GAAGACCAAGGAGTGGTGGATCGGCACATAGAGGTCAGCCTGAAAGGAAAGGGGTCTTC 3773
Qy 997 ProValSerPheValHisPheIleAlaAsp 1006
Db 3774 CCAATGCTCTTCGTGCACATTTCTGTGCAGAC 3803

RESULT 9
ADS34277
ID ADS34277 standard; DNA; 5947 BP.
XX
AC ADS34277;
XX
DT 02-DEC-2004 (first entry)
XX
DE POSH protein associated DNA #31.
XX
KW ds; gene; cytostatic; nootropic; neuroprotective; antiparkinsonian;
KW anticonvulsant; antiviral; neuroleptic; central nervous system;
KW POSH polypeptide; POSH-associated protein; POSH-AP; HERPUD1;
KW Ubiquitin ligase; antiviral agent; anti-apoptotic agent;
KW anti-cancer agent; secretory pathway trafficking inhibitor;
KW neurological disorder progression disorder; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; schizophrenia;
KW Niemann-Pick's disease.
XX
OS Homo sapiens.
XX
PN WO2004078130-A2.
XX
PD 16-SEP-2004.
XX
02-MAR-2004; 2004WO-US006308.
03-MAR-2003; 2003US-0451437P.
05-MAR-2003; 2003US-0452284P.
19-MAR-2003; 2003US-0455760P.
20-MAR-2003; 2003US-0456640P.
03-APR-2003; 2003US-0460526P.
04-APR-2003; 2003US-0460792P.
21-APR-2003; 2003US-0464285P.
09-MAY-2003; 2003US-0469462P.
15-MAY-2003; 2003US-0471378P.
20-MAY-2003; 2003US-0472327P.
30-MAY-2003; 2003US-0474706P.
03-JUN-2003; 2003US-0475825P.
17-JUN-2003; 2003US-0479317P.
19-JUN-2003; 2003US-0480215P.
19-JUN-2003; 2003US-0480376P.
08-AUG-2003; 2003US-0493860P.
28-AUG-2003; 2003US-0498634P.
16-SEP-2003; 2003US-0503931P.
10-NOV-2003; 2003WO-US035712.
05-FEB-2004; 2004WO-US003600.
02-MAR-2004; 2004US-0549896P.
XX
(PROT-) PROTEOLOGICS INC.
XX
Taglicht DN, Alroy I, Reiss Y, Year L, Ben-Avraham D, Tuvia S;
PI Greener T;
XX
WPI; 2004-662346/64.
XX
Isolated, purified or recombinant complex, useful for identifying an
PT antiviral, anti-apoptotic or anti-cancer, comprising POSH polypeptide and
PT POSH-associated protein (POSH-AP).

```


[illegible]

Db	2814	GAAC	CTGCCCCCAAAACACACAGCTGGGGGACCTGCCACCCAAACCACTCTCAGACTTA	2877
Qy	888	ProGln	LyS-----	890
Db	2874	CCTCC	AAACCAACAGATGAAGGACCTGCCCCCAACACCAAGCTCTGAGGTCACACTGAAG	2933
Qy	891	-----	-----LySProAlaProGlyThrAsp	898
Db	2934	AAATCC	CAGACTGGAGATGTCTCACCAAGGCTCAGCAACCTCTGAGGTCACACTGAAG	2993
Qy	899	SerThr	Pro-----LeuThrAsnLysGlyGlnProArgGlyProValAspLeuSerAla	916
Db	2994	TCAC	CCCATTTGGATCTATCCCAAAATGTGCAGTCCAGAGACGCCATCCAAAAGCAAGCA	3053
Qy	917	ThrGlu	AlaLeuGlyProLeuSerAsnAlaMetValLeuGlnProProAlaProMetPro	936
Db	3054	TCTGA	AGACTCCCAACGCTCAGCCCTACTCTGCCA---GAGACGCCGTACCACTGCC	3110
Qy	937	ArgLys	SerGlnAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysVal	956
Db	3111	AGAAAA	TAATACGGGGGAAAAATAAAGTGAGCGAGTGAAGACCATTTATGATGCTCCAG	3170
Qy	957	AlaAsp	AsnProAspGluLeuThrPheSerGluGlyAspVallelleValleAspGlyGlu	976
Db	3171	GCAG	ACAACGATGACGAGCTCACATTATCGAGGGGAGAAGTGATTATCGTCACAGGGGA	3230
Qy	977	GluAsp	GlnGluTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPhe	996
Db	3231	GAGAC	CAAGAGTGTGGATTGGCCACATCGAAGGACAGCCTGAAGGAAGGGGGTCTTT	3290
Qy	997	ProVal	SerPheValHisPheIleAlaAsp 1006	
Db	3291	CCAG	TGTCCTTTGTTTCATATCCTGCTGAC 3320	
RESULT	10			
ID	ADS10088			
ID	ADN10088	standard; DNA; 4997 BP.		
AC	ADS10088;			
XX				
DT	16-DEC-2004	(first entry)		
DE	Human therapeutic DNA - SEQ ID 325.			
KW	antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnerary;			
KW	inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;			
KW	aplastic anaemia; cancer; wound healing; gene therapy; ds; gene.			
OS	Homo sapiens.			
XX				
PN	WO2004080148-A2.			
XX				
PD	23-SEP-2004.			
XX				
PF	30-SEP-2003; 2003WO-US030720.			
XX				
PR	02-OCT-2002; 2002US-0416186P.			
XX				
PA	(NUVE-) NUVELO INC.			
XX				
PI	Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;			
PI	Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;			
XX				
DR	WPI; 2004-668857/65.			
XX				
DR	P-PSDB; ADS10772.			
XX				
PT	New polynucleotide, useful in preparing a composition for diagnosing or			
PT	treating inflammatory, neurodegenerative or stem cell disorders, e.g.,			
PT	aplastic anemia or cancer for promoting wound healing.			
XX				
FS	Claim 1; SEQ ID NO 325; 718pp; English.			
XX				
CC	The invention relates to a novel isolated polynucleotide and the encoded			

100

CC polypeptide. The molecules of the invention demonstrate antiinflammatory,
 CC neuroprotective, antianaemic, cytostatic and vulnerary activities and may
 CC be useful in preparing a composition for diagnosing or treating
 CC inflammatory, haematopoietic, immune, neurodegenerative or stem cell
 CC disorders, such as aplastic anaemia or cancer, as well as for promoting
 CC wound healing. The molecules may also be utilised during gene therapy
 CC procedures. The current sequence is that of a human therapeutic DNA of
 CC the invention. The current sequence is not shown explicitly within the
 CC specification but can be accessed from the WIPO web-site.

XX Sequence 4997 BP; 1539 A; 1189 C; 1025 G; 1244 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6 25e-104	Length:	4997
Score:	2556.50	Matches:	546
Percent Similarity:	63.97%	Conservative:	118
Best Local Similarity:	52.60%	Mismatches:	212
Query Match:	48.48%	Indels:	162
DB:	13	Gaps:	23

US-09-914-042-1 (1-1006) x ADS10088 (1-4997)

Qy	95	GlySerAlaPheLeuLysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsn	114
Db	2	GGCACCGGTTTGTCAAGTTTCTACTCTTACAAAGGAACGTCCACACTGCTGAAAAAT	61
Qy	115	LeulleGlnAsnMetAsnAsnIleIleSerPheProLeuAspSerLeuLysGlyAsp	134
Db	62	CTGCTCCAGGGTTTGAGCCACAATGTGATCTTACCTTGGATCTTTGTTAAAGGAGAC	121
Qy	135	LeuLysGlyValLysGly-AspLeuLysLysPheAspLysAlaTyrLysAspTyrG1	154
Db	122	CTAAGGGAGTCAAGGATACA-----	143
Qy	154	uThrLysIleThrLysIleGluLysGluLysGluHisAlaLysLeuHisGlyMetI1	174
Db	144	-----AAATTTGAGAAAGAAAGAGAGAGCAGCAGCAAAACACATGGGATGAT	190
Qy	174	eArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGluuArgArgPhePh	194
Db	191	CCGCACAGAGATAACAGGAGCTGAGATTGCGGAAGAAATGGGAAGAGGAGCGCTCTT	250
Qy	194	eGlnLeuGlnMetCysGluTyrLeuLysValAsnGluIleLysIleLysGlyVa	214
Db	251	TCAGTCCCAATGTGTGATATCTCTTAAAGTTAATGAATCAAGACCAAAAGGGTGT	310
Qy	214	lAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAspG1	234
Db	311	GGATCTGTCGAGAAATCTTATAAAGTATTACCATGCACAGTGCAATTTCTTCAAGATGG	370
Qy	234	YLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHisTh	254
Db	371	CTTGAAACACAGCTGATAAGTTTGAACAGTACATTGAAAAAATGGCTGCTGATTATATAA	430
Qy	254	rIleLysGlnAlaGlnAspGluLysArgGlnLeuIleGlnLeuArgAspIleLeuLys	274
Db	431	TATAAAACAGACCCAGGAGTGAAGAAAGAAACAGCTAACTGCCACTCCGAGACTTAATAA	490
Qy	274	sSerAlaLeuGlnValGluLysGlu-----AspSerGlnIleArgGlnSerTh	291
Db	491	ATCTCTCTTCAACTGGATCAGAAAGAAATCTAGGAGAGATTTCTCAGACCGGCAAGGA--	548
Qy	291	rAlaTyrSerLeuHisGlnProGlnGlyAsnLysGluHisGlyThrGluuArgAsnGlyse	311
Db	549	-GGATACAGCATGATCAGCTCCAGGCAATAAGGAATATGCGAGTGAAGAAAGGGGTA	607
Qy	311	rLeuTyrLysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerVally	331
Db	608	CCTGTAAGAAGTGTGCGGATCCGGAAGTATGGCAGAGGGAAGTGTTCAGTCAA	667
Qy	331	sAsnGlyPheLeuThrIleSerHisGlyThrAlaAsnArgProAlaLysLeuAsnLe	351
Db	668	GAATGGGATTCTAACCAATCTACATGCCCATCTAACAGGCACACCCNAGTTCAACT	727

Qy	351	uLeuThrCysGlnValLysThrAsnProGluGluLysLysCysPheAspLeuLysSerHi	371
Db	728	TCTCACTCGCAAGTAAAACTTANTGCGAAGACAAAAATCTTTTGACCTGATATCACA	787
Qy	371	sAspArgThrTyrHisPheGlnAlaGluAspGluGlnCysGlnIleTrpMetSerVa	391
Db	788	TAATAGAACATATCACTTTTTCAGGCAGAGATGAGCAGGATTTATGTAGCATGGATACGT	847
Qy	391	lLeuGlnAsnSerLysGluLysGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrG1	411
Db	848	ATTGACAAATAGCAAAAGAGAGCCCTAACCATGCGCTTCCGTGGAGAGAGAGTGGCGG	907
Qy	411	YGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetTh	431
Db	908	AGAGAACAGCGTG---GAAGACCTGCACAAAGCCATTTATGAGGATGTCACAGCGGCTCCC	964
Qy	431	rGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLe	451
Db	965	AGGGAATGACATTTGCTGCGATTGCTCATCAGAACCCACCTGGCTTTCAACCAACTT	1024
Qy	451	uGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSe	471
Db	1025	GGGTATTTTGACCTGTATAGAAATGCTTGGCATCCATAGGAAATGGGGGTTCATATTTTC	1084
Qy	471	rProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAs	491
Db	1085	TCGCATTCAGTCTTTGGAACCTAGACAAATTAGGAACCTTCTGAACCTTCTGCTGGCCAGNA	1144
Qy	491	nIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerVa	511
Db	1145	TGTAGGAACAATAGTTTATATGATATTTATGGAAGCAAAATTTTACCAGC---CCCTCAC	1201
Qy	511	lLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrI1	531
Db	1202	AAAAACCCACCCCTTCAAGTGTATGACTGTACGAAAAAGAAATATATCATCTGCAAAAGTATGT	1261
Qy	531	eGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLeuCy	551
Db	1262	AGATCATAGGTTTTCAAGGAAGACCTGTTCAACTTCATCAGCTAACTAAATGAATGCT	1321
Qy	551	sGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAs	571
Db	1322	TGAGGCCATCAATCCAGGGATTTACTTGCCTACTTAATCAAGTCTATGCAAGGGGTAGA	1381
Qy	571	pLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisdu	591
Db	1382	GCTAATGGAACCACTG---CTGGAACCTGGGCGAGAGCTTGGGAGACAGCCCTTCACCT	1438
Qy	591	uAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSe	611
Db	1439	TGCGCTCGAATCGAGATCAGACATCTCTCAATTTGGTTGACTTCTCTTGTACAAACTG	1498
Qy	611	rGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysLeuTh	631
Db	1499	TGGGAACCTGATTAAGCAGACGGCCCTGGGAAACACAGTCTTACACTACTGTAGTATGTA	1558
Qy	631	rAspAsnAlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAs	651
Db	1559	CAGTAAACCTGAGTGTGTTGAAGCTTTTCTCAGGAGCAAGCCCACTGTGGATATAGTTAA	1618
Qy	651	nGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluG1	671
Db	1619	CCAGGCTGGAGAAACTGCCCTAGACATAGCAAGAGACTAAAGACTCCAGTGTGTAAGA	1678
Qy	671	uLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTr	691
Db	1679	TCTGCTTCCAGCGGTAAATCTGGAAGTTCAATCCACACCGTCCACGTAGAAATAGGTG	1738
Qy	691	pArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnPr	711
Db	1739	GAATCTTCACAGGAGGAGATAGATGAGCGGATGATGTGATGATGATGATGATGATGATG	1798

Qy 711 oSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLe 731
 Db 1799 TATCAAGAAAGAGCGCTCACCAGACCTTCAGAGCTTCCGCACTCTCCACAGCTCTCCCC 1858
 Qy 731 uGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArg 751
 Db 1859 CCAGGAC-----AAGCTGGCACTCCAGGATTACAGACTCCAGGGACAAACAGCG 1909
 Qy 751 gAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSerGly---- 769
 Db 1910 GCTC-----TCCTATGGAGCTTCACCAACAGATCTT 1942
 Qy 770 ----SerProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro 788
 Db 1943 CGTTTCCACAAGCAGACTCGCCACATACCA---ACCACGAGGCTCCCCCTCTGCC 1999
 Qy 788 oProArgAsnValGlyLys-----ValG1 796
 Db 2000 TCCTAGGAACCGCGGAAGGTCCCACTGGCCCACTTCAACACTCCCTCTAAGCACCA 2059
 Qy 796 nThrAlaSerSerAlaAsnThrLeuTrpLysThrAsnSerValSerValAspGlyGlyse 816
 Db 2060 GACTCTAGTGGCAGCTCCACTATCCAGAAGAGGCTCTCTCCCCCACCACCGGA-- 2117
 Qy 816 xArgGlnArgSerSerAspProProAlaValHisProProLeuProProLeuArg-- 835
 Db 2118 -CACAAGAGAACCTATATCCGACCTCCAGCCCACTACCTCTATGGGGCCCCCAACAAGG 2176
 Qy 835 ----- 835
 Db 2177 CGCATGTCCTTGGGGTAAGATGGGGTCCATCTCTTCAAGTAAGACTACAAACAGTT 2236
 Qy 836 -----ValThrSerThrAsnPro---Le 842
 Db 2237 TGAGGGACTATCCAGCAGTCGAGCACCAGTTCTGCAAGAGACTGCCCTTGGCCCAAGAGT 2296
 Qy 842 uThrProThrProProProValAla-----LysThr----- 853
 Db 2297 TCTTCTAAACTACTCTCAGAAAGTGGCACTAAGGAAACAGATCATCTCTCCCTAGACAA 2356
 Qy 854 -----ProSerValMetGluAlaLeuSerGln-----ProSerLy 865
 Db 2357 AGCCACCATCCGCCCGCGGAAATCTTTCAGAAATCATCAGTTGGCAGAGTTGCCACAAA 2416
 Qy 865 sProAlaPro-----ProGlyIleSe 872
 Db 2417 GCCACCACTGGAGACTGCCCCCAAGCCACAGACTGCGCCCAAGCCCAAAATGG 2476
 Qy 872 rGlnIleArgProProPro-----LeuProPr 881
 Db 2477 AGATTGCGGCTAAGCCAGGAGAACTGCCCCCAACACAGCTGGGGGACCTGCCACC 2536
 Qy 881 oGlnProPro---SerArgLeuProGlnLys----- 890
 Db 2537 CAACCCCAACTCTCAGACTTACTTCCCAACCAACAGATGAAGGACCTGCCCCCAAC 2596
 Qy 891 -----Ly 891
 Db 2597 ACAGCTGGAGACCTGTAGCAAAATCCAGACTGGAGATCTCTACCAAGGCTCAGCA 2656
 Qy 891 sProAlaProGlyThrAspLysSerThrPro-----LeuThrAsnLysGlyGlnProAr 909
 Db 2657 ACCCTCTGAGTGCACACTGAAGTCACACCCATTCGATCTATCCCAAAATGTGAGTCCAG 2716
 Qy 909 gGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAsnAlaMetValLe 929
 Db 2717 AGCGCATCCAAAGCAAGCATCTGAAGACTCCCAACGACTCCAGCTACTCTGCGCA-- 2774
 Qy 929 uGlnProProAlaProMetProArgLysSerGlnAlaThrLysLeuLysProLysArgVa 949
 Db 2775 -GAGACGCGCTACACTGCTCCCAAGAAATCAATACGGGGGAAATAAAGTGGCGGAGT 2833
 Qy 949 lLysAlaLeuTyrAsnCysValAlaAspAsnProAspGluLeuThrPheSerGluGlyAs 969

Db 2834 GAAGACCATTTATGACTGCCAGGCAGACACGATGACGAGCTCATTTCATCGAGGAGA 2893
 Qy 969 pValIleIleValAspGlyGluGluAspGlnGluTrpIleGlyHisIleAspGlyAs 989
 Db 2894 AGTCATTTATCTCAGGGGAGAGGACGAGAGTGGTGGATTGGCCACATCGAAGGACA 2953
 Qy 989 pProGlyArgLysGlyAlaPheProValSerPheValHisPheIleAlaAsp 1006
 Db 2954 GCCTGAAGGAAGGGGTCTTTCAGTGTCTTGTTCATATCTCTGTCTGAC 3005
 RESULT 11
 AAIS9224
 ID AAIS9224 standard; cDNA; 5033 BP.
 XX
 AC AAIS9224;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1427.
 DE
 KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Aeundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 WIPI; 2001-442253/47.
 DR P-FSDB; AAM40068.
 DR
 XX
 Novel nucleic acids and polypeptides, useful for treating disorders such
 as central nervous system injuries.
 PT
 XX
 Claim 1; SEQ ID NO 1427; 10078pp; English.
 PS
 XX
 The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 encoded polypeptides (AAM38642-AAM42213) with nontropic,
 immunosuppressant and cytostatic activity. The polynucleotides are useful
 in gene therapy. A composition containing a polypeptide or polynucleotide
 of the invention may be used to treat diseases of the peripheral nervous
 system, such as peripheral nervous injuries, peripheral neuropathy and
 localised neuropathies and central nervous system diseases, such as
 Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 utilisation of the activities such as: Immune system suppression,
 Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 assays for receptor activity, arthritis and inflammation, leukaemias and

CC useful for identifying an agent that modulates an activity of a POSH
 CC polypeptide or POSH-AP, for identifying an antiviral agent, an anti-
 CC apoptotic agent, an anti-cancer agent, an agent that inhibits trafficking
 CC of a protein through the secretory pathway, an agent that inhibits the
 CC progression of a neurological disorder, an agent that modulates a POSH
 CC function, an agent that modulates a HERPUD1 function. The methods can be
 CC used for treating a viral infection, for inhibiting an activity of a POSH
 CC -AP in a cell, for treating a POSH-associated disease in a subject. The
 CC POSH-associated disease is viral infection, POSH-associated cancer or
 CC POSH-associated neurological disorder. The methods are useful for
 CC treating or preventing POSH-associated neurological disorder in a subject
 CC e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC schizophrenia, Niemann-Pick's disease. This sequence corresponds to a
 CC nucleic acid of the invention.

XX SQ Sequence 4870 BP; 1509 A; 1159 C; 997 G; 1205 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6,36e-100	Length:	4870
Score:	2465.50	Matches:	520
Percent Similarity:	64.62%	Conservative:	110
Best Local Similarity:	53.33%	Mismatches:	204
Query Match:	46.76%	Indels:	141
DB:	13	Gaps:	21

US-09-914-042-1 (1-1006) x ADS34272 (1-4870)

Qy	154	GlutThrIleThrLysIleGluLysGluLysLysGluHisAlaLysLeuHisGlyMet	173
Db	1	GAGACAAAGTTTACAAAATTCAGAAAGAGAAAGAGAGACGCGCAAAACACATGGGATG	60
Qy	174	IleArgThrGluIleSerGlyAlaGluIleAlaGluMetGluLysGluArgArgPhe	193
Db	61	ATCGGCACAGAGATAACAGAGAGCTGAGATTGCGGAAGAAATGGAGAGAAAGCGCGCTC	120
Qy	194	PheGlnLeuGlnMetCysGluTyrLeuLeuLysValAsnGluIleLysIleLysGly	213
Db	121	TTTCAGCTCCAAATGTGTGAATATCTCATTAAAGTTAATGAATCAACACCAAAAGGGT	180
Qy	214	ValAspLeuLeuGlnAsnLeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAsp	233
Db	181	GTGGATCTGCTGCAGAACTTATATAAGATTATACCATGCACAGTGCAATTTCTTTCAAGAT	240
Qy	234	GlyLeuLysAlaValGluSerLeuLysProSerIleGluThrLeuSerThrAspLeuHis	253
Db	241	GGCTTGAAGAACAGCTGATAAGTTGAACAGTACATTGAAACACTGGCTGCTGATTATAT	300
Qy	254	ThrIleLysGlnAlaGlnAspGluArgGlnLeuIleGlnLeuArgAspIleLeu	273
Db	301	AATATAAACAGACCCAGGATGAAGAAAGAAACAGCTAATGCACTCCGAGACTTAATA	360
Qy	274	LysSerAlaLeuGlnValGluLysGluAspSerGlnIleArgGlnSerThrAlaTyr	293
Db	361	AAATCCTCTCTCAACTGGATCAGAAAGAAAGATTCTCAGAGCGCGGAGGA---GGATAC	417
Qy	294	SerLeuHisGlnProGlnClyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyr	313
Db	418	AGCATGCATACGCTCCAGGGCAATGAAGAAATATGGCAGTGAAGAAAGGGGTACCTGCTA	477
Qy	314	LysLysSerAspGlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGly	333
Db	478	AAGAAAAGTACGGGATCCGGAAGTATGGCAGAGAGAGGAAGTGTTCAGTCAAGAAATGGG	537
Qy	334	PheLeuThrIleSerHisGlyThrAlaIleArgProAlaLysLeuAsnLeuLeuThr	353
Db	538	ATTCTGACCATCTCACATGCCACATCTAACAGGCAACCCAGGCAAGTGTGAACCTTCTACC	597
Qy	354	CysGlnValLysThrAsnProGluLysLysCysPheAspLeuIleSerHisAspArg	373
Db	598	TGCCAAGTAACACCTTAATGCCGAAGCAAAAAATCTTTTGACCTGATATCATATAATAGA	657
Qy	374	ThrTyrHisPheGlnAlaGluAspGluGlnGluCysGlnIleTrpMetSerValLeuGln	393

Db	658	ACATATCACTTTTCAGGCAGAGATGAGCAGGATTATTGTAGCATGGATATCATGATTGACA	717
Qy	394	AsnSerLysGluGluAlaLeuAsnAsnAlaPheLysGlyAspAspAsnThrGlyGluAsn	413
Db	718	AATAGCAAGAAGAGAGCCCTAAACCATGGCTTCGTGGAGAGCAGAGTGGCGGAGAGAAC	777
Qy	414	AsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArgMetThrGlyAsn	433
Db	778	AGCTG---GAAGACCTGACAAAAGCCATTATTAGGATGTCACGGCTCCAGGGAT	834
Qy	434	AspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThrAsnLeuGlyIle	453
Db	835	GACATTTGCTGCGATTGTGGCTCATCAGAACCCACCTGGCTTCAACCAACTGGGTATT	894
Qy	454	LeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHisTyrSerProMet	473
Db	895	TTGACCTGTATAGAAATGTTCTGGCATCATAGGAAATGGGGTTTCATATCTCTCGCAAT	954
Qy	474	GlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAlaLysAsnIleGly	493
Db	955	CAGTCTTTGGAACTAGACAAATTAGGAACCTTCTGAACTCTTCTGGCCAGAAATGTAGA	1014
Qy	494	AsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAspSerValLysPro	513
Db	1015	AACAATAGTTTTTAATGATATTTATGGAAGCAAAATTTACCAGC---CCCTCACCAAAACCC	1071
Qy	514	AsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLysTyrIleGluArg	533
Db	1072	ACCCCTTTCAAGTGTATGACTGTGTACGAAAGAAATATATCACTGCAAAAGTATGTAGATCAT	1131
Qy	534	ArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSerLysCysGluAla	553
Db	1132	AGGTTTTCAAGGAAGACCTGTTCAACTTCATCAGCTAAACTTAATGAATTCCTTGAGGCC	1191
Qy	554	ValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThr	573
Db	1192	ATCAAAATCCAGGGATTTACTTTGCACATAATCAAGTCTATGCAAGAGGGGTAGAGCTAATG	1251
Qy	574	GluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeuHisLeuAlaVal	593
Db	1252	GAGCACCTG---CTGGAACCTGGCGAGAGCTTGGGAGAGACAGCCCTTCACCTTCGCGCTC	1308
Qy	594	ArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGlnAsnSerGlyAsn	613
Db	1309	CGAACTGCAGATCAGACATCTCTCCATTTGTTGCTTCTTGTACAAACCTGTGGGAAC	1368
Qy	614	LeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCysLeuThrAspAsn	633
Db	1369	CTGGATAAGCAGACGGCCCTGGGAACACACAGTTCTACACTACTGTAGTATGTACAGTAAA	1428
Qy	634	AlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIleAlaAsnGluSer	653
Db	1429	CCTGAGTGTTTGAAGCTTTTGCTCAGGAGCAAGCCCACTGTGGATATAGTTAAACAGGCT	1488
Qy	654	GlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCysGluGluLeuLeu	673
Db	1489	GGAGAACTGCCCTAGACATAGCAAGAGACTAAAGCTACCCAGTGTGAAGATCTGCTT	1548
Qy	674	ThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyrGluTrpArgLeu	693
Db	1549	TCCAGGCTAAATCTGGAAAGTTCAATCCACACGCTCCACGCTAGATAATATGAGTGGAACTT	1608
Qy	694	LeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeuGlnProSerGlu	713
Db	1609	CGACAGGAGGAGATAGATGAGAGCGATGATCTGGATGACAAACCAAGCCCTTATCAAG	1668
Qy	714	AsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSer	733
Db	1669	AAAGAGCGCTCACCCAGACCTCAGAGCTTCTGCCACTCTCCAGCATCTCCCCCAGGAC	1728
Qy	734	AsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLysGlnArgAlaPhe	753
Db	1729	-----AAGCTTGGCACTGCCAGGATTCAGCACTTCCAGGGGCAAAACAGCGGCTC---	1776

CC progression of a neurological disorder, an agent that modulates a POSH
 CC function, an agent that modulates a HERPUD1 function. The methods can be
 CC used for treating a viral infection, for inhibiting an activity of a POSH
 CC -AP in a cell, for treating a POSH-associated disease in a subject. The
 CC POSH-associated disease is viral infection, POSH-associated cancer or
 CC POSH-associated neurological disorder. The methods are useful for
 CC treating or preventing POSH-associated neurological disorder in a subject
 CC e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC schizophrenia, Niemann-Pick's disease. This sequence corresponds to a
 CC nucleic acid of the invention.

XX
 SQ Sequence 5475 BP; 1680 A; 1264 C; 1086 G; 1445 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.02e-97 Length: 5475
 Score: 2410.00 Matches: 511
 Percent Similarity: 64.15% Conservative: 110
 Best Local Similarity: 52.79% Mismatches: 203
 Query Match: 45.70% Indels: 144
 DB: 13 Gaps: 22

US-09-914-042-1 (1-1006) x ADS34273 (1-5475)

Qy 164 LysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSerGlyAlaGluIle 183
 Db 2 AAAGAGAGCAGCGAAACACATGGGATGATCCGACAGAGATTAACAGGAGCTGAGATT 61
 Qy 184 AlaGluMetGluLysGluArgPhePheGlnLeuGlnMetCysGluTyrLeuLeu 203
 Db 62 GCGGAAGAAATGGAGAGAAAGGCGCTTTTCAGCTCCAAATCGTGAATATCTCAT 121
 Qy 204 LysValaGluIleLysIleLysGlyValAspLeuLeuGlnAsnLeuIleLysTyr 223
 Db 122 AAAGTTAATGAATCAAGACCAAAAGGTTGGATCTGTCGCAAAATCTTATAAGTAT 181
 Qy 224 PheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGluSerLeuLysPro 243
 Db 182 TACCATGACAGTGCATTTCTTCAGATGGCTTGAANAACAGCTGATAAGTTGNAACAG 241
 Qy 244 SerLeuGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGlnAspGluArg 263
 Db 242 TACATTGAAAACATGGCTGCTGATTTATATATAATAAAGGCTTCTTCAACTGGATCAGAAAG 301
 Qy 264 ArgGlnLeuIleGlnLeuArgAspIleLysSerAlaLeuGlnValGluGlnLysGlu 283
 Db 302 AAACAGCTAACTGCACCTCGAGACTTAATAAAATCTCTTCAACTGGATCAGAAAGAA 361
 Qy 284 -----AspSerGlnIleArgGlnSerThrAlaTyrSerLeuHisGlnProGlnGly 300
 Db 362 TCTAGGAGAGATTCTCAGAGCGGCAAGGA---GGATACAGATGCATCAGCTCCAGGCG 418
 Qy 301 AsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAspGlyIleArg 320
 Db 419 AATAAGGAATATGGCAGTGAAAGAGGGGTACCTGCTTAAAGAAAGTGAACGGGATCCGG 478
 Qy 321 LysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIleSerHisGly 340
 Db 479 AAAGTATGCGAGAGAGAGAGTGTTCAGTCAAGAAATGGGATTTCTAACCATCTCACATGCC 538
 Qy 341 ThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLysThrAsnPro 360
 Db 539 ACATCTACAGGCAACAGCCAGTTGAACCTTCTCACCTGCCAAGTAAACCTTAACTGCC 598
 Qy 361 GluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPheGlnAlaGlu 380
 Db 599 GAAGACAAAAAATCTTTTGACCTGATATACATAATAGAAATACATACATCTTTCAGGAGAA 658
 Qy 381 AspGluGlnGluCysGlnIleTrpMetSerValLeuGlnAsnSerLysGluGluAlaLeu 400
 Db 659 GATGACGAGGATATGTGATGATGATATATGATGATGATGATGATGATGATGATGATGAT 718
 Qy 401 AsnAlaAlaPheLysGlyAspAsnThrGlyGluAsnAsnIleValGlnGlnLeuThr 420

Db 719 ACCATGCGCTCCGTGAGAGACAGTGTGGGAGAGAAACAGCCTG---GAAGACCTGACA 775
 Qy 421 LysGluIleLeuSerGluValGlnArgMetThrGlyAsnAspValCysCysAspCysGly 440
 Db 776 AAACCCATTATTGAGGATGTCACCGCTCCAGGGAATGACATTTGCTGGATTGTGGC 835
 Qy 441 AlaProAspProThrTrpLeuSerThrAsnLeuGlyIleLeuThrCysIleGluCysSer 460
 Db 836 TCATCAGAACCCACTGCTTTCAACCACTGGGTATTTTACCTGCTGATAGAATGTTCT 895
 Qy 461 GlyIleHisArgGluLeuGlyValHisTyrSerProMetGlnSerLeuThrLeuAspVal 480
 Db 896 GGCATCCATAGGAAATGGGGTTTCATATTCTCGCATTCAGTCTTTTGAACCTAGACAA 955
 Qy 481 LeuGlyThrSerGluLeuLeuAlaLysAsnIleGlyAsnAlaGlyPheAsnGluIle 500
 Db 956 TTAGAAATCTCTGAACCTTCTGTCGCAAGAAATGTAGAAACAATAGTTTAAATGATATT 1015
 Qy 501 MetGluCysCysLeuProAlaGluAspSerValLysProAsnProGlySerAspMetAsn 520
 Db 1016 ATGAAGCAAAATTTACCAGC---CCCTCACCAAAACCCACCCTTCAAGTGATATGACT 1072
 Qy 521 AlaArgLysAspTyrIleThrAlaLysTyrIleGluArgArgTyrAlaArgLysLysHis 540
 Db 1073 GTACGAAAGAAATATATACCTGCAAGATGATGTAGATCATAGTTTTCAGGAAAGACCTGT 1132
 Qy 541 AlaAspAsnAlaAlaLysLeuHisSerLeuCysGluAlaValLysThrArgAspIlePhe 560
 Db 1133 TCAACTTCATCAGCTAAACTAAATGTAATGCTTGAGGCCATCAAAATCCAGGATTTACTT 1192
 Qy 561 GlyLeuLeuGlnAlaTyrAlaAspGlyValAspLeuThrGluLysIleProLeuAlaAsn 580
 Db 1193 GCATTAATCAAGTCTATGCAAGGGGTAGAGCTAATGGAAACCAGCTG---CTGAACCT 1249
 Qy 581 GlyHisGluProAspGluThrAlaLeuHisLeuAlaValArgSerValAspArgThrSer 600
 Db 1250 GGGCAGGAGCTTGGGAGACAGCCCTTCACTTCCGCTCCGAACCTGCAGATCAGACATCT 1309
 Qy 601 LeuHisIleValAspPheLeuValGlnAsnSerGlyAsnLeuAspLysGlnThrGlyLys 620
 Db 1310 CTCCATTGGTGTGACTTCTTGTACAAACCTGTGGGAACCTGGATTAAGCACACGCGCTG 1369
 Qy 621 GlySerThrAlaLeuHisTyrCysLeuThrAspAsnAlaGluCysLeuLysLeuLeu 640
 Db 1370 GGAACACAGCTTCTACACTACTGTAGTATGTACAGTAAACCTGAGTGTGTGAAGCTTTG 1429
 Qy 641 LeuArgGlyLysAlaSerIleGluIleAlaAsnGluSerGlyGluThrProLeuAspIle 660
 Db 1430 CTCAGGAGCAAGCCACTGTGGATATAGTTAAACAGGCTGGAGAACTGCCCTAGACATA 1489
 Qy 661 AlaLysArgLeuLysHisGluHisCysGluGluLeuLeuThrGlnAlaLeuSerGlyArg 680
 Db 1490 GCAAGAGACTTAAAGACTACCCAGTGTGAAGATCTGCTTCCAGGCTTAAATCTGGAAG 1549
 Qy 681 PheAsnSerHisValHisValGluTyrGluTrpArgLeuLeuHisGluAspLeuAspGlu 700
 Db 1550 TTCAATCCACAGCTCCACGTAGATATAGTGGAAATCTTCCACAGGAGGAGATAGATGAG 1609
 Qy 701 SerAspAspMetAspGluLysLeuGlnProSerGluAsnArgArgGluAspArgPro 720
 Db 1610 AGCGATGATGATCTGGATGACAAACCAAGCCCTATCAAGAAAGAGCGCTCACCAGACT 1669
 Qy 721 IleSerPheTyrGlnLeuGlySerAsnGlnLeuGlnSerAsnAlaValSerLeuAlaArg 740
 Db 1670 CAGAGCTTCTGCCACTCTCCAGCATCTCCCCCAGGAC-----AAGCTGGCAGCTG 1720
 Qy 741 AspAlaAlaLeuAlaLysGluLysGlnArgAlaPheMetProSerIleLeuGlnAsn 760
 Db 1721 CCAGATTACAGACTCCNAGGGACAAACAGCGGCTC----- 1756
 Qy 761 GluThrTyrGlyAlaLeuLeuSerGly-----SerProProAlaGlnProAla 777
 Db 1757 ---TCCTATGGAGCTTCCACCAACAGATCTTCGTTTCCCAAGACACAGACTCGGCCACA 1813

Alignment Scores:

Pred. No.: 2,01e-83 Length: 3812
 Score: 2089.50 Matches: 455
 Percent Similarity: 59.20% Conservative: 153
 Best Local Similarity: 44.30% Mismatches: 275
 Query Match: 39.63% Indels: 145
 DB: 10 Gaps: 19

US-09-914-042-1 (1-1006) x ADB62827 (1-3812)

Qy 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrIleVala 20
 Db 47 ATGCCGAGAGTTCAGCGTGCCTGAGTTCCTGGCGGTCCACCGCGGAGGACCTCAGCTCC 106
 Qy 21 ProThr----AlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAla 39
 Db 107 CCGGTGGGGCGCGCGCTTCGCCGCCAAGATGCCCGGTACCCGAGGGCGCGCTGGCG 166
 Qy 40 IleGluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysLysSerVallys 59
 Db 167 CGGGAGGAGACTTGGAAAGGAGACCAAGCCATCTTCGACAGAAATAAAGAGCTGTGCG 226
 Qy 60 AlaIleAsnSerSerGlyLeuAlaHisValGluAsnGluGluInTyrThrGlnAlaLeu 79
 Db 227 GCAATCCATAGTCCGGCTTGGCCATGTGGAGATGAAGACAGTCCGAGAGGCCGTG 286
 Qy 80 GluLysPheGlyAsnCyseValCysArgAspAspProAspLeuGlySerAlaPheLeu 99
 Db 287 GAATCTTTAGGCAACAGCACCTGTCCAGAACACGCCATGAGCTTCCACAGGCTTCTA 346
 Qy 100 LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMet 119
 Db 347 AACTTGGCGGTGTTCCCGCGAGGTGCTGCGCTTTCAGNACTGATTCAGAACTTG 406
 Qy 120 AsnAsnIleIleSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyVallys 139
 Db 407 AACAACTTGTCTTTTCCCCCTGCACAGTCTGATGAAGGGCGACGTGAGGGACGGTCA 466
 Qy 140 GlyAspLeuLysLysPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLys 159
 Db 467 CAGGATTCCAAAAAACAGCTGGAGAGGCGATGGAAGGACTATGAAGCCAAAATGGCCAA 526
 Qy 160 IleGluLysGluLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSer 179
 Db 527 CTGGAGAGGAG--CGGATCCGGCCAGGTGACAGGAGGATC----- 568
 Qy 180 GlyAlaGluIleAlaGluMetGluLysGluArgArgPhePheGlnLeuGlnMetCys 199
 Db 569 CTGGGAGGTGGCCAGACATGCAGAGAGCGCGCATCTTCAGCTGCACATGTGT 628
 Qy 200 GluTyrLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsn 219
 Db 629 GAGTATCTGCTCAAAACCGGGGAGAGCCAGATGAAGCAAGGTCTCTTCAGAGC 688
 Qy 220 LeuIleLysTyrPheHisAlaGlnCysAsnPhePheGlnAspGlyLeuLysAlaValGlu 239
 Db 689 CTCATCAAGTCTTCACGCCCCAGCACAACTTTTTCAGATGGCTGGAAGGCTGCCAG 748
 Qy 240 SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln 259
 Db 749 AGCTGTTCCTTCATCGAAGCTGGCGGCTCAGTACATGCATCCATCAGGCCAG 808
 Qy 260 AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279
 Db 809 GAGGACGAGCTACAGAAGCTGACCCAGCTCCGGACTCCCTCCGAGGAGCACATGCACTT 868
 Qy 280 GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln 297
 Db 869 GAGACGACGAGGAACACTGAGCCGGAAGAACTCAGGATGTGGCTATAGCATCCACCAG 928
 Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysLysSerAsp 317
 Db 929 CACCAAGGCAACAGCAGCTTTGGGAGCGGAGAAAGTGGGCTTTCTATACAAGAAAGTGAC 988

Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
 Db 989 GGAATTCGAGAGTCTGCGAGAAAAGGAAGTGTGGAGTCAAGTATGGCTGCTGACCAATC 1048
 Qy 338 SerHisGlyThrAlaAsnArgProProAlaLysLeuAsnLeuLeuThrCysGlnVallys 357
 Db 1049 TCACACAGCAGATAAACCCGCCCGGTGAAGCTGACCCCTGCTGACGTGCCAAGTGAGG 1108
 Qy 358 ThrAsnProGluGluLysLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
 Db 1109 CCAAAACCTGAGGAGAAAAGTGTGACCTGTGACCCACCAACACCGACGTACCACTTT 1168
 Qy 378 GlnAlaGluAspGluGlnLysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
 Db 1169 CAGGACAGGAGCAGACGAGTGTGAGCGTGGGTGTGAGTGTTCAGAAACAGACAGAGAC 1228
 Qy 398 GluAlaLeuAsnAlaPheLysGlyAsp-----AspAsn 409
 Db 1229 GAAGCCCTGAGCAGCGCTTCTCGGGGAGCCACAGCGTGGCCCGGGGTCTTGGGGGTCC 1288
 Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleIleSerGluValGlnArg 429
 Db 1289 GCCGGCCATGATGGGAGCGCACGACCTCAAAAGCTGTCTCATCGCGAGGTGAAGAGC 1348
 Qy 430 MetThrGlyAsnAspValCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449
 Db 1349 AGGCTGGGAATGCCAGTGTGCGACTGCGGGGTGCAGACCCACCGTGTGACACCC 1408
 Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
 Db 1409 AACCTGGCGGTGCTCACCTGCATCCAGTGTCTCGGGCGTCCACCGCGAACTATGCGCGC 1468
 Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
 Db 1469 TTTTTCGCGCATGCACTCACTTGCACCTTGCACCTGCGGCCCTCCGAGTGTGTGTCGCC 1528
 Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysLeuProAlaGluAsp 509
 Db 1529 TTGAACATGGGAACACAGAGCTTCAATGAGTCTATGAGGCGCCAGCTACCTCCACCGC 1588
 Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
 Db 1589 GGCCTTAAACCTCAGCTGAGAGTGACATGGCGACCCCGCAGCGACTACATATGSCCAAG 1648
 Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
 Db 1649 TATGTGGAGCATAGTGTTCACCGCGG-----TGACACACCTGAGCGCTCAGCGA 1696
 Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
 Db 1697 CTCTGGACAGCATTTGCAACAGGACCTCTCTGCGTACTGGAGGCTTTGCCNATGGG 1756
 Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
 Db 1757 CAGGACTTTGACAGCGCTCCAGGCGCTGATGCACAGGACCTGGAAGAACTCGTCTTG 1816
 Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
 Db 1817 CATTTGGGTGTCAAAAGTCGCAACAGCGCTTCCCTCGCTCTGCTGCTGATCATCAG 1876
 Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
 Db 1877 AACGGTGTCACTGGATGCCAAGGCTCTGACGGGAACACGGCTCTGCATCTACCGACA 1936
 Qy 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuArgGlyLysAlaSerIleGluIle 649
 Db 1937 CTCTACAACACGACCGCGTGTCAAGCTGTCTGCTGAAGGGGAGAGCTTTGTTGGCACA 1996
 Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
 Db 1997 GTAAATGAAGCAGCGGAGACAGCTCTGGACATAGCCAGGAAGAACACCAAGAGGTGT 2056

QY 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2057 GAGGAGCTGCTGGAGCAGCCCGGAGCTTTTGGCTTCCCTACATGTCGATAC 2116
QY 690 GluTTPArgLeuLeuHisGluAspLeuAspGluSerAspMetAspGluValLeu 709
Db 2117 TCCTGGTAATTCACAGAGCTGGCTGACAGTGAGAGGATGAGGAGAGAG--- 2173
QY 710 GlnProSerGluAsnArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2173 ----- 2173
QY 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2174 ----- 2174
QY 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
Db 2204 CACTGGCCAGTGGGAGGCTGGACATCAGCAACAGACCTATGAGACTGTGCCAGCCTG 2263
QY 769 GlySerProProProAlaGlnProAlaAlaProSerThrThrSerAlaProProLeuPro 788
Db 2264 GGAGCA-----GCCACCCTCAGGCGAGAGTGAGGAGCTGCCCGCCCTTGCCA 2314
QY 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsn 808
Db 2315 GTCAAAAC-----TCTTCTCGGACTTTG----- 2338
QY 809 SerValSerValAspGlyClySerArgGlnArgSer-----Ser-SerAspProPr 825
Db 2339 -----GTCAAGGGTGTCCAGACATGCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2389
QY 825 AlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2390 CTGAGTTGAGAGCCCTGAGACCCCTGAGAGCTGGGAGTCCAG----- 2435
QY 845 rProProProProAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2436 -CCTCTCTCCAGTCTGATGAGCCCT-----TGGACCTGGGATCCAGCCAA 2485
QY 865 aProAlaProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905
Db 2486 GCGCCACCACTCTGAGAGGGCTCCGAGAGCCCGGAGGACCTCCAGACCCAGCCTG 2545
QY 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905
Db 2546 ACAT-----CCGGACCCCTTCGGAGATGATC----- 2576
QY 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2577 -----TCCCGCTCAGAT---TCAGCTCCGAGAGCCTCGCTCTATCGCGCG 2620
QY 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerGln 940
Db 2621 GGGCGCGAGCCCTGAGATGCTCCCTCAGCAGG-CAGCCTCTGCCCAAGAGAGCT 2679
QY 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2680 GCCG----- 2683
QY 960 aAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnGln 980
Db 2683 ----- 2683
QY 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2684 -----GTTGCATCACTGAGAGAGATGGCTCAAGAGACTGGAGTCTCCCAAGAGTTC 2736
QY 1000 eValHisPheIleAlaAsp 1006
Db 2737 TGTGCAACTTTTGCAAGAC 2755

ADK81946
ID ADK81946 standard; DNA; 4050 BP.
XX AC ADK81946;
DT 06-MAY-2004 (first entry)
XX DE Hepatocellular carcinoma up-regulated gene DDEF1L.
KW ds; gene; cytostatic; antisense gene therapy; up-regulation;
KW hepatocellular carcinoma; antisense.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 46..2757
FT /*tag= a
FT /product= "DDEF1 protein"
FT /gene= "ddef1l"
XX PN WO2003027322-A2.
XX PD 03-APR-2003.
XX PF 25-SEP-2002; 2002WO-JP009873.
XX PR 25-SEP-2001; 2001US-0324261P.
XX PR 23-AUG-2002; 2002CA-02399569.
XX PR (UITY) UNIV TOKYO.
XX PA (ONCO-) ONCOTHERAPY SCI INC.
XX PI Nakamura Y, Furukawa Y;
XX WPI; 2003-371927/35.
XX P-PSDB; ADK81947.
XX PT New isolated nucleic acid, polypeptide encoded by it and antibody against
XX the polypeptide, for detecting hepatocellular carcinoma in a subject.
XX PS Claim 1; SEQ ID NO 1; 89pp; English.
XX CC The invention relates to isolated nucleic acids comprising genes that are
CC up-regulated in hepatocellular carcinoma cells; or comprising strands
CC that hybridize under high stringent conditions to the genes or a
CC complement. The nucleic acids, polypeptides encoded by them, or
CC antibodies against the polypeptide are used to detect hepatocellular
CC carcinoma in a subject. Antisense oligonucleotides to the nucleic acids
CC are used for inhibiting growth of hepatocellular carcinomas. This
CC sequence corresponds to the DDEF1 gene which encodes a protein with
CC similarity to a centaurin protein and contains an Arf GTPase-activating
CC protein domain and 2 ankyrin repeats.
XX SQ Sequence 4050 BP; 910 A; 1163 C; 1162 G; 815 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.13e-83 Length: 4050
Score: 2089.50 Matches: 455
Percent Similarity: 59.20% Conservative: 153
Best Local Similarity: 44.30% Mismatches: 275
Query Match: 39.63% Indels: 145
DB: 10 Gaps: 19

US-09-914-042-1 (1-1006) x ADK81946 (1-4050)
QY 1 MetProAspGlnIleSerValSerGluPheValAlaGluThrHisGluAspTyrLysAla 20
Db 46 ATGCGGAGCAGTTCAGCGTCCCGAGTTCCTGCGCGTCACCGCGAGGACCTCAGCTCC 105
QY 21 ProThr---AlaSerSerPheThrThrArgThrAlaGlnCysArgAsnThrValAlaAla 39
Db 106 CCGGCTGGGGCGCGCGCTTCGCCGCCAAGATGCCCGGTACCGAGGGGCGCGCTGGCG 165


```
Qy 40 IleGluGluAlaLeuAspValAspArgMetValLeuTyrLysMetLysSerValLys 59
Db 166 CGGAGGAGATCTTGGAGGAGACCAAGCCATCTGACAGATATAAGAGGCTGTGCGG 225
Qy 60 AlaleAsnSerSerGlyLeuAlaHisValGluAsnGluGluGlnTyrThrGlnAlaLeu 79
Db 226 GCAATCCATAGCTCCGGCTTGGCCATGTGGAGATGAAGAGCAGTACCGAGAGGCCGTG 285
Qy 80 GluLysPheGlyGlyAsnCyValCyArgAspAspProAspLeuGlySerAlaPheLeu 99
Db 286 GAATCTTTAGCAACACAGCACCTGTCGCCAGACAGCCATGAGCTCCACAGGCTTCCTA 345
Qy 100 LysPheSerValPheThrLysGluLeuThrAlaLeuPheLysAsnLeuIleGlnAsnMet 119
Db 346 AACTTGGCGGTGTTCCCGGAGGTTGCTGGCTCTTCAAGAACTGATTCAGAACTTG 405
Qy 120 AsnAsnIleLeuSerPheProLeuAspSerLeuLeuLysGlyAspLeuLysGlyValLys 139
Db 406 AACAACTGTTCTTCCCTCGACAGTCTGATGAAGGGGCGAGCTGAGGGACGGTCCA 465
Qy 140 GlyAspLeuLysLysPheAspLysAlaTrpLysAspTyrGluThrLysIleThrLys 159
Db 466 CAGGATCCAAAAACACGCTGAGAGAGGATGGAAGGACTATGAAGCCANAATGGCCAAAG 525
Qy 160 IleGluLysGluLysLysGluHisAlaLysLeuHisGlyMetIleArgThrGluIleSer 179
Db 526 CTGGNAGAGGAG--CGGATCGGCCAGGTGACAGGAGGATC----- 567
Qy 180 GlyAlaGluIleAlaGluGluMetGluLysGluArgArgPheGlnLeuGlnMetCys 199
Db 568 CTGGGAGGTGGCCCGACGACATGCAGAGAGCGCGCATCTTCAGCTGCACATGTGT 627
Qy 200 GluTyrLeuLeuLysValAsnGluIleLysLysLysGlyValAspLeuLeuGlnAsn 219
Db 628 GAGTATCTGCTCAAAAGCCGGGAGGAGCCAGATGAAGCAAGTCTCTGACTTCTTCAGAGC 687
Qy 220 LeuIleLysTyrPheHisAlaGlnCysAsnPheGlnAspGlyLeuLysAlaValGlu 239
Db 688 CTCATCAAGTCTTCCAGCCCGACGACAACTTTTCCAAAGATGGTGGAGGCTGCCCAG 747
Qy 240 SerLeuLysProSerIleGluThrLeuSerThrAspLeuHisThrIleLysGlnAlaGln 259
Db 748 AGCTGTGTTCCCTTCATCGAAGAGTGGCGGCTCAGTACATGCATCCATCAGGCCAG 807
Qy 260 AspGluGluArgArgGlnLeuIleGlnLeuArgAspIleLeuLysSerAlaLeuGlnVal 279
Db 808 GAGGACGAGCTACAGAAGTACACGAGCTCCGGGACTCCCTCCGAGGAGCACTGCAGCTT 867
Qy 280 GluGlnLysGluAspSerGlnIleArgGlnSerThr-----AlaTyrSerLeuHisGln 297
Db 868 GAGACGACAGAGGAACACCTGAGCCGGAAGAACTCAGGATGTGGCTATAGCATCCACCAG 927
Qy 298 ProGlnGlyAsnLysGluHisGlyThrGluArgAsnGlySerLeuTyrLysSerAsp 317
Db 928 CACCAAGGCAACACGAGTTTGGGACGGAGAAAGTGGGCTTCTATACAAGAAAGTGCAC 987
Qy 318 GlyIleArgLysValTrpGlnLysArgLysCysSerValLysAsnGlyPheLeuThrIle 337
Db 988 GGAATTCGAAGAGTCTGCGAGAAAAGGAAGTGTGGAGTCAAGTATGGCTGTGACCATC 1047
Qy 338 SerHisGlyThrAlaAsnArgProAlaLysLeuAsnLeuLeuThrCysGlnValLys 357
Db 1048 TCACACACGACGATAAAACCGGCCCCCGGTGAAGCTGACCCCTGCTGACGTGCCAGTGAGG 1107
Qy 358 ThrAsnProGluGluLysCysPheAspLeuIleSerHisAspArgThrTyrHisPhe 377
Db 1108 CCAAAACCTCGAGGAGAAAAGTCTTCGACCTGGTGAACCCACCAACCGGACGTACCACCTT 1167
Qy 378 GlnAlaGluAspGluGlnCysGlnIleTrpMetSerValLeuGlnAsnSerLysGlu 397
Db 1168 CAGGACGAGGACGACGACGAGTGTGAGCGCTGGGTGTGAGTGTTCGAAACAGCAGGAGAC 1227
Qy 398 GluAlaLeuAsnAsnAlaPheLysGlyAsp-----AspAsn 409
```

```
Db 1228 GAAGCCCTGACGAGCGCTTCTCGGGAGGCCAGCGCTGCGCGGGTCTCTGGGGTCC 1287
Qy 410 ThrGlyGluAsnAsnIleValGlnGluLeuThrLysGluIleIleSerGluValGlnArg 429
Db 1288 GCCGCCCATGATGGGAGCCGACAGCCTCAAAAGCTGTCTCATCGCGAGGTGAAGAGC 1347
Qy 430 MetThrGlyAsnAspValCysCysAspCysGlyAlaProAspProThrTrpLeuSerThr 449
Db 1348 AGGCTGGGAATAGCCAGTGTCTGCGACTGCGGGGCTGCAGACCCACACGTGCTCAGACC 1407
Qy 450 AsnLeuGlyIleLeuThrCysIleGluCysSerGlyIleHisArgGluLeuGlyValHis 469
Db 1408 AACCTGGGCGTGTCTCACTGCTGCTCGGGCGTCCACCCGGAACCTGGCGTGC 1467
Qy 470 TyrSerProMetGlnSerLeuThrLeuAspValLeuGlyThrSerGluLeuLeuAla 489
Db 1468 TTTTGGCGCATGTCAGTCACTCACCTTGGACCTGTGGGCCCTCCGAGTTGTTGTGCGCC 1527
Qy 490 LysAsnIleGlyAsnAlaGlyPheAsnGluIleMetGluCysCysValLeuProAlaGluAsp 509
Db 1528 TTGAACATGGGAACAACGAGCTTCATGAGGTCTGAGGCCAGCTACCTCACCGGC 1587
Qy 510 SerValLysProAsnProGlySerAspMetAsnAlaArgLysAspTyrIleThrAlaLys 529
Db 1588 GGCCCTAAACCTCAGCTGAGAGTGACATGGGCACCCCGCAGGAGCTACATTTATGGCCAAG 1647
Qy 530 TyrIleGluArgArgTyrAlaArgLysLysHisAlaAspAsnAlaAlaLysLeuHisSer 549
Db 1648 TATGTGAGCATAGTGTTCACGCGCG-----TGACACCTCGAGCCTCAGCGA 1695
Qy 550 LeuCysGluAlaValLysThrArgAspIlePheGlyLeuLeuGlnAlaTyrAlaAspGly 569
Db 1696 CTCTGGACAGCAATTCGCAACAGGAGCTCTCTGCTGGTACTGGAGGCTTTGCCAATGGG 1755
Qy 570 ValAspLeuThrGluLysIleProLeuAlaAsnGlyHisGluProAspGluThrAlaLeu 589
Db 1756 CAGGACTTTGGACAGCCGCTGCCAGGCTGTATGCACAGGCACCTGAAGAACTCGTCTTG 1815
Qy 590 HisLeuAlaValArgSerValAspArgThrSerLeuHisIleValAspPheLeuValGln 609
Db 1816 CATTTGGGTGTCAAGTCGCCAACAGGCTTCCCTGCTCTCTGGTGGATTTTCATCCAG 1875
Qy 610 AsnSerGlyAsnLeuAspLysGlnThrGlyLysGlySerThrAlaLeuHisTyrCysCys 629
Db 1876 AACGCTGTCTACCTGGATGCCAGCTGCTGACGGGAACACGCTCTGCATATCGCAGCA 1935
Qy 630 LeuThrAspAsnAlaGluCysLeuLysLeuLeuLeuArgGlyLysAlaSerIleGluIle 649
Db 1936 CTCTACAACACGACCGGCTGCTCAAGCTGCTGAAGGGGAGAGCTTTGTTGGGCACA 1995
Qy 650 AlaAsnGluSerGlyGluThrProLeuAspIleAlaLysArgLeuLysHisGluHisCys 669
Db 1996 GTAAATGAAGCAGCGACAGCTCTGACATATGCCAGGAAGAAGACCAACAAGAGGTGT 2055
Qy 670 GluGluLeuLeuThrGlnAlaLeuSerGlyArgPheAsnSerHisValHisValGluTyr 689
Db 2056 GAGGAGCTGTCTGGACAGGCCCGCGGGACCTTTGCTTCCCTCTACATGTGGACTAC 2115
Qy 690 GluTrpArgLeuLeuHisGluAspLeuAspGluSerAspAspMetAspGluLysLeu 709
Db 2116 TCCTGGGTAAATTTCCACAGAGCCCTGCTCAGAGTGAGGAGGATGAGGAAGAAG--- 2172
Qy 710 GlnProSerGluAsnArgArgGluAspArgProIleSerPheTyrGlnLeuGlySerAsn 729
Db 2172 ----- 2172
Qy 730 GlnLeuGlnSerAsnAlaValSerLeuAlaArgAspAlaAlaAsnLeuAlaLysGluLys 749
Db 2173 -----CGCTGCTTGTGAAGCTCCCGGCCGAGCT 2202
Qy 750 GlnArgAlaPheMetProSerIleLeuGlnAsnGluThrTyrGlyAlaLeuLeuSer--- 768
```


Db 2203 CACTGGCCAGTGGGAGCTGGACATCAGCAACAGACCTATGAGACTGTGCCAGCCTG 2262
Qy 769 GlySerProProAlaGlnProAlaAlaProSerThrThrSerAlaProLeuPro 788
Db 2263 GGAGCA-----GCCACCCCTCAGGGCGAGAGTGTCCCGCCCTTGCCA 2313
Qy 789 ProArgAsnValGlyLysValGlnThrAlaSerSerAlaAsnThrLeuTrpLysThrAsn 808
Db 2314 GTCAAAAAC-----TCTCTCGGACTTTG----- 2337
Qy 809 SerValSerValAspGlySerArgGlnArgSer-----Ser-SerAspProPr 825
Db 2338 -----GTCCAGGTTCGCAAGATGCCAGTGGAGATCGTTCTGAAGTCTCCAGC 2388
Qy 825 oAlaValHisProProLeuProProLeuArgValThrSerThrAsnProLeuThrProTh 845
Db 2389 CTGAGTTCCAGAGGCCCTCAGACCCCTGAGAGCCTGGCGAGTCCAG----- 2434
Qy 845 rProProProValAlaLysThrProSerValMetGluAlaLeuSerGlnProSerLys 865
Db 2435 -CCTCCTCTCCAGTCTGATGAGCCCT-----TGGAACCTGGGGATCCAGCCAA 2484
Qy 865 sProAlaProProGlyLysSerGlnIleArgProProProLeuProGlnProProSe 885
Db 2485 GCCCCACCAACTCTGAAGAGGGCTCCGAGAGCCCCAGGACCTCCAGACCCAGCCTG 2544
Qy 885 rArgLeuProGlnLysLysProAlaProGlyThrAspLysSerThrProLeuThrAsnLys 905
Db 2545 ACAT-----CCGGGACACCCCTTCGGAGATGATACC----- 2575
Qy 905 sGlyGlnProArgGlyProValAspLeuSerAlaThrGluAlaLeuGlyProLeuSerAs 925
Db 2576 -----TCCCGCTCAGAT---TCAGCTCCGAGAGCACTCGCTCCTATCGCGG 2619
Qy 925 nAla-----MetValLeuGlnProProAlaProMetProArgLysSerG1 940
Db 2620 GGGGGCGGAGCCCTGAAGATGGTCCCTCAGCCAGG-CAGCCTCTGCCAGAGGAACGT 2678
Qy 940 nAlaThrLysLeuLysProLysArgValLysAlaLeuTyrAsnCysValAlaAspAsnPr 960
Db 2679 GCGG----- 2682
Qy 960 oAspGluLeuThrPheSerGluGlyAspValIleValAspGlyGluGluAspGlnG1 980
Db 2682 ----- 2682
Qy 980 uTrpTrpIleGlyHisIleAspGlyAspProGlyArgLysGlyAlaPheProValSerPh 1000
Db 2683 -----GTTGGCATCACTGAAGAGAGATGGCTCAAGGACTGGGAGTCTCCAGCAAGTTC 2735
Qy 1000 eValHisPheIleAlaAsp 1006
Db 2736 TGTGCACTTTTGCAGAC 2754

Search completed: August 4, 2005, 08:58:12
Job time : 1406 secs

THIS PAGE BLANK (USPTO)